Equilibrium and Kinetic Folding Pathways of a TIM Barrel with a Funneled Energy Landscape

John M. Finke and José N. Onuchic
The Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92093-0374

ABSTRACT The role of native contact topology in the folding of a TIM barrel model based on the α-subunit of tryptophan synthase (αTS) from *Salmonella typhimurium* (Protein Data Bank structure 1BKS) was studied using both equilibrium and kinetic simulations. Equilibrium simulations of αTS reveal the population of two intermediate ensembles, I_1 and I_2, during unfolding/refolding at the folding temperature, $T_f = 335$ K. Equilibrium intermediate I_1 demonstrates discrete structure in regions α_0-β_6 whereas intermediate I_2 is a loose ensemble of states with N-terminal structure varying from at least β_1-β_3 (denoted I_{2A}) to α_0-β_4 at most (denoted I_{2B}). The structures of I_1 and I_2 match well with the two intermediate states detected in equilibrium folding experiments of *Escherichia coli* αTS. Kinetic folding simulations of αTS reveal the sequential population of four intermediate ensembles, I_{120Q}, I_{200Q}, I_{300Q}, and I_{360Q}, during refolding. Kinetic intermediates I_{120Q}, I_{200Q}, and I_{300Q} are highly similar to equilibrium αTS intermediates I_{2A}, I_{2B}, and I_1, respectively, consistent with kinetic experiments on αTS from *E. coli*. A small population (~10%) of kinetic trajectories are trapped in the I_{120Q} intermediate ensemble and require a slow and complete unfolding step to properly refold. Both the on-pathway and off-pathway I_{120Q} intermediates show structure in β_1-β_3, which is also strikingly consistent with kinetic folding experiments of αTS. In the off-pathway intermediate I_{120Q}, helix α_2 is wrapped in a nonnative chiral arrangement around strand β_3, sterically preventing the subsequent folding step between β_3 and β_4. These results demonstrate the success of combining kinetic and equilibrium simulations of minimalist protein models to explore TIM barrel folding and the folding of other large proteins.

INTRODUCTION

Understanding the fundamental physics of protein folding is a goal of both experimentalists and theoreticians. Guided by landscape theory (1–8), an understanding of the fundamental principles of protein folding has recently advanced due to the development of small fast-folding peptide systems that are tractable to study by all-atom simulation and theory (9–16) and minimalist simulation models that can effectively sample the dynamics of larger protein systems (17–23). Although these research efforts are increasing our understanding of protein folding, many challenges remain.

Many recent studies have applied these minimalist models in molecular dynamics studies of the folding of intermediate length proteins (17,19–21). These studies have played a crucial step in validating landscape theory, because protein models with funneled energy landscapes show remarkable success in predicting experimentally determined structural and thermodynamic properties of protein folding pathways. However, it remains to be seen to what degree larger proteins will also obey the principle of minimal frustration. Furthermore, although minimalist models have proved highly advantageous in sampling smaller proteins, it is unclear as to how well they will sample folding states of larger proteins. It is necessary to use this basic same model in simulations to determine whether larger proteins can also be accurately described, thereby demonstrating that large proteins also have a low level of energetic frustration. Simulations of large proteins using minimalist molecular models will highlight some of the possible future challenges and computational shortcuts in all-atom simulations of large systems.

This study uses folding simulations of tryptophan synthase, a member of the TIM (triose phosphate isomerase) barrel structural family, to determine whether the energy landscape of this large protein is funneled as has been shown for smaller proteins (19,21,24). The TIM barrel is an excellent large protein structure to investigate with minimalist Go-models because it is the most ubiquitous protein family (10% of Protein Data Bank (PDB)). A fundamental understanding of TIM barrel folding could be applied to thousands of proteins across every genome.

Folding experiments of monomeric TIM barrels have been conducted on yeast (γTIM), rabbit muscle (rTIM), and *Trypanosoma brucei* (tbTIM) triose phosphate isomerase, the α-subunit of *Escherichia coli* tryptophan synthase (αTS), *Sulfolobus solfataricus* indole-3-glycerol phosphate synthase (sIGPS), *E. coli* indole-3-glycerol phosphate synthase (eIGPS), *E. coli* phosphoribosylanthranilate isomerase (PRAI), and rabbit muscle aldolase (25–43).

The stability and folding pathways of TIM barrel proteins have been the subject of discussion because the basic structure comprises repeating $\beta\alpha$-units in a circular arrangement connecting the N- and C-termini, with the most common structure having eight $\beta\alpha$-units. With respect to stability, it may appear that all eight units would be required for folding and activity because all eight strands are required

Submitted January 5, 2005, and accepted for publication April 11, 2005.

Address reprint requests to José N. Onuchic, Center of Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093. Tel.: 858-534-7067; E-mail: jonuchic@ucsd.edu.

0006-3495/05/07/488/18 $2.00
doi: 10.1529/biophysj.105.059147

to make contacts between the N- and C-terminal strands. However, truncation mutants of αTS have demonstrated that stable structure can exist in the N-terminal region of the protein without the C-terminal residues (40).

In addition, protein folding experiments have revealed that different TIM barrel proteins do not always fold in the same manner, although they typically fold in steps involving contiguous $\beta\alpha$-units. Intermediates are typically observed in equilibrium unfolding experiments of most TIM barrels (26–28,31,32,35,39,43,44), with the possible exception of rTIM (36). Kinetic folding studies of all TIM barrels studied demonstrate multiphasic folding pathways and are consistent with folding intermediates (30–32,35,36,41,42).

Despite this common observation, the properties of folding intermediates, both equilibrium and kinetic, can be quite different between different TIM barrel proteins. The equilibrium folding pathway of the α-subunit of tryptophan synthase (αTS) from *E. Coli* involves an initial folding intermediate I_2 within regions α_0–α_4 (38,40) followed by an intermediate I_1 comprising regions α_0–$\beta_6 + \beta_7$ (37,40,45). The kinetic folding pathway of αTS shows early structure in regions α_0–$\beta_6 + \beta_7$ (30,45), as was found with intermediate I_1 in the equilibrium experiments. A similar equilibrium folding pathway was found in γTIM where initial folding initiates with intermediate I_2 comprising regions β_2–β_4 followed by an intermediate I_1 comprising regions α_1–β_6 (27). In contrast, no intermediates are observed in equilibrium unfolding of rTIM although an intermediate comprising the C-terminal regions β_5–α_8 is observed in kinetic refolding experiments (36). Also, a dialysis refolding experiment indicates that the folding pathway of rabbit muscle aldolase populates two intermediates with noncontinuous structural units α_0, β_4–β_8 (I_1) and α_0, $\beta_4\alpha_4$, α_5, $\alpha_6\beta_7$ (I_2) (34).

The fact that members of the TIM barrel structural family do not rigorously conserve their folding pathways suggests that overall fold is not sufficient to explain TIM barrel folding. Thus, the folding mechanism must be reflected in other subtler properties. One possible indicator of TIM barrel folding differences may be the slight differences in the contact topology due to different position and lengths of the α-helices and β-strands of each TIM barrel family member. Although contact topology certainly plays a role in folding, a possible additional determinant is differential energetic weighting from different amino acid types. Also, nonnative contacts, proline isomerization, and disulphide formation can play an important role in protein folding (42,46,47), although these phenomena are not explicitly included in models of this study for simplicity.

The primary question addressed in this work is whether the information provided from a contact map of the TIM barrel protein αTS is sufficient to accurately describe its folding pathway in a minimalist protein model, which would indicate that contact topology is the primary determinant of αTS folding (2,17,48). This would also support the hypothesis that the energy landscape of αTS is highly funneled to the native state. This work addresses this hypothesis using kinetic and thermodynamic simulations of minimalist model for αTS. Previous simulations of αTS without a funneled energy landscape also appear to successfully qualitatively capture the dominant intermediate in the αTS kinetic folding pathway (23). This previous study suggests that a non-Go-model can determine the TIM barrel folding pathways and ultimately discriminate the final native state. This study addresses whether energetically funneled Go-models can capture the experimentally determined equilibrium and kinetic pathway of αTS.

MATERIALS AND METHODS

Molecular dynamics

Molecular dynamics (MD) simulations were carried out using AMBER 6 software, compiled on a Linux platform, employing the sander_classic program as an integrator for initial energy minimization and subsequent molecular dynamics (49). The following describes the AMBER sander_classic molecular dynamics parameters used in this study. The specific parameter values are listed in parentheses. The time step was 0.001 ps ($DT = 0.001$). Translational and rotational motion was removed at the beginning of each run and every 1000 time steps thereafter ($NTCM = 1$, $NSCM = 1000$, $NDFMIN = 0$). Initial velocities were randomly selected ($INIT = 3$, $IG =$ random). If the absolute value of the velocity of any atom exceeded 500 Å per time step, velocities are scaled such that the absolute value of the velocity of that atom = 500 Å per time step ($VLIMIT = 500$). Temperature was maintained with external bath using the method of Berendsen (50) with a coupling constant of 0.2 ps ($NTT = 5$, $TAUTP = 0.2$, $TAUTS = 0.2$). If the simulation temperature T_{sim} exceeds the average temperature T by >10 K, velocities are scaled such that $T_{sim} = T$. SHAKE was not used. Although no electrostatics were involved in the molecular dynamics, a default constant dielectric was used ($IDIEL = 1$) with a default delectric constant of 1 ($DIELC = 1$). The particle mesh Ewald method was not used ($IEWALD = 0$). During each integration step, interactions between all atom pairs were calculated and this contact pair list was updated only once at the beginning of the simulation ($CUT = 9999$, $NSNB = 9999$). No periodic boundary and pressure regulation were used ($NTB = 0$, $NTP = 0$). Structures and energies were saved every 1.5 ps ($NTPR = 1500$, $NTWR = 1500$, $NTWX = 1500$, $NTWV = 1500$, $NTWE = 1500$).

Go-model

To model αTS, each amino acid in the α-subunit of *Salmonella typhimurium* tryptophan synthase is approximated with its single backbone C_α atom from the PDB file 1BKS, as shown in Fig. 1. All atoms not found in the α-subunit of tryptophan synthase were not included in the model. It should be noted that the model is based on the αTS structure from *S. typhimurium* and assumes that the structure of the α-subunit will be the same in the absence of the β-subunit. Ultimately, comparison is made between simulations and experiments on the isolated αTS subunit from *E. coli* (29–31,37–42,45,51). Nonetheless, it is logical to conclude that the isolated αTS structure and folding pathway of *S. typhimurium* will be similar to that of *E. coli* due to the nearly identical sequence conservation. For this reason, references to ''αTS'' apply to the α-subunit of tryptophan synthase of either *S. typhimurium* or *E. coli*, unless otherwise specified.

The overall potential energy for a given protein conformation is given by Eq. 1:

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{LJ} + E_{rep}. \quad (1)$$

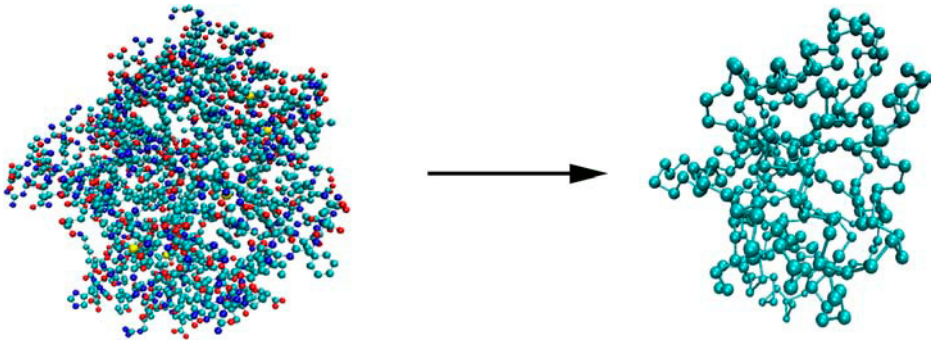


FIGURE 1 Conversion of the PDB coordinates of 1BKS to a C_α model of αTS.

Consistent with the original Go-model (52), the minimum energy of each energy term is obtained when the protein is in the native folded state.

For covalent bond distance terms,

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2}\varepsilon_r(r - r_0)^2, \tag{2}$$

where $\varepsilon_r = 100$ kcal mol^{-1} Å^{-2} is the bond energy, r is the bond distance in the simulation, and r_0 is the native $C_\alpha C_\alpha$ bond distance in the reduced C_α PDB structure, summed over all bonds in the reduced C_α PDB structure. The minimum energy $C_\alpha C_\alpha$ bond distance, r_0, in the disordered region of αTS (residues 178–191) is set to 3.81 Å.

For the bond angle term,

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{1}{2}\varepsilon_\theta(\theta - \theta_0)^2, \tag{3}$$

where $\varepsilon_\theta = 20$ kcal mol^{-1} degree^{-2} is the bond angle energy, θ is the bond angle in the simulation, and θ_0 is the $C_\alpha C_\alpha C_\alpha$ native bond angle in the reduced C_α PDB structure, summed over all bond angles in the reduced C_α PDB structure. The minimum energy $C_\alpha C_\alpha C_\alpha$ bond angle, θ_0, in the disordered region of αTS (residues 178–191) is set to 109.5°.

For dihedral energies,

$$E_{\text{dihdedral}} = \sum_{\text{dihedrals}} \Big[\varepsilon_\phi^1[1 - \cos(\phi - \phi_0)] + \varepsilon_\phi^2[1 - \cos(3(\phi - \phi_0))]\Big], \tag{4}$$

where $\varepsilon_\phi^1 = 0.8$ kcal/mol, $\varepsilon_\phi^2 = 0.4$ kcal/mol, ϕ is the dihedral angle in the simulation, and ϕ_0 is the $C_\alpha C_\alpha C_\alpha C_\alpha$ native dihedral angle in the reduced C_α PDB structure, summed over all dihedral angles in the reduced C_α PDB structure. The energies of $C_\alpha C_\alpha C_\alpha C_\alpha$ dihedrals in the disordered region of αTS (residues 178–191) are set to $\varepsilon_\phi^1 = 0$ kcal/mol and $\varepsilon_\phi^2 = 0$ kcal/mol, effectively producing a flexible linker between αTS residues 177 and 192.

In the Go-model, two C_α atoms in a protein were selected as attractive if they are separated by four or more residues and are indicated to be in contact using contacts of structural units analysis (53). Each attractive C_α-C_α contact is described by an attractive Lennard-Jones potential

$$E_{\text{LJ}} = \sum_{|i-j|>3} \varepsilon_{\text{LJ}}\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right], \tag{5}$$

where $\varepsilon_{\text{LJ}} = 0.8$ kcal/mol is the contact energy, σ_{ij} is the native distance between the two contact atoms, i and j, given from the crystal structure, and r_{ij} is the distance between the two contact atoms, i and j, determined for a given iteration of the simulation.

If any two atoms are not determined to be attractive or fall within three residues of each other (i,i + 3), then their interaction is defined by a repulsive term

$$E_{\text{rep}} = \sum_{i,j} \varepsilon_{\text{rep}}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}, \tag{6}$$

where $\varepsilon_{\text{rep}} = 0.8$ kcal/mol is the repulsive energy, σ_{ij} is half the hard-sphere distance between two repulsive atoms i and j (1.9 Å), and r_{ij} is the distance between the two repulsive atoms, i and j, determined for a given iteration of the simulation.

The parameters used in the Go-model are shown in Table 1. These parameters were selected because they had produced an accurate folding temperature and stability of chymotrypsin inhibitor 2 in previous work (24).

Thermodynamics

The initial structure used for MD studies of αTS was obtained from simulated annealing, using the 1BKS PDB coordinates as an initial structure (Fig. 1). For αTS, MD simulations were run for 420 ns at 330 K, 270 ns at 335 K, 1300 ns at 340 K, and 580 ns at 345 K. For each structure sampled throughout the simulations, the number of native contacts (Q) formed was calculated where each native contact was determined to be formed if it falls within 1.5 times the native distance. Thermodynamic quantities, such as free energy (G), energy (E), and entropy (S), are determined using all simulation temperatures simultaneously with the weighted histogram analysis method (WHAM) algorithm (54). Using WHAM, the potential mean force (PMF) is plotted versus the number of native contacts in the protein (Q) at a temperature, T, where a range of Q values are sampled using Eq. 7.

$$PMF(Q = X) = -k_B T \log\left(\frac{\sum_{i}^{Q=X} W_T * n(E_i) * e^{-\frac{E_i}{k_B T}}}{\sum_{i}^{Q=\text{all}} W_T * n(E_i) * e^{-\frac{E_i}{k_B T}}}\right). \tag{7}$$

In Eq. 7, k_B is the Boltzman constant, $n(E_i)$ is the density of states in the simulation with the indicated value of Q, W_T is the WHAM-converged numerical weight for iterations in the trajectory with temperature T, $Q = X$

TABLE 1 Parameters of αTS Go-model

Parameter	Energy ε
Bonds	(kcal/mol)
$C_\alpha C_\alpha$	100
Angles	(kcal/mol)
$C_\alpha C_\alpha C_\alpha$	20
Dihedrals	(kcal/mol)
$C_\alpha C_\alpha C_\alpha C_\alpha$	0.8 (ε_ϕ^1)
–	0.4 (ε_ϕ^2)
10–12 Contacts	(kcal/mol)
$C_\alpha C_\alpha$	0.8 (i,i + 4)

denotes all simulation configurations with X native contacts, and $Q = ALL$ denotes all simulation configurations. Although PMF is not a direct measure of free energy, differences in PMF are equivalent to the difference in free energy (ΔG). For example, the free-energy difference (ΔG_{Q1Q2}) between two discrete values of Q, i.e., folded $Q_1 = 400$ versus unfolded $Q_2 = 25$, can be calculated with Eq. 8:

$$\Delta G_{Q1Q2} = PMF(Q_1) - PMF(Q_2). \quad (8)$$

Kinetics

For kinetic refolding simulations, 60 kinetic trajectories are collected to obtain statistically significant reaction rate measurements. The initial unfolded coordinates of each refolding trajectory are obtained from the final structure of a short simulation at 999 K of a randomly determined length (500–1500 ps) and random initial velocities, followed by 1 ns at 373 K and random initial velocities. For each refolding trajectory, these initial 373 K coordinates are subjected to 300 K and random initial velocities and followed for a minimum of 30 ns (30×10^6 time steps). Although 30 ns was a sufficient number of computational steps to refold most αTS trajectories, trajectories that did not refold were simulated further until the native ensemble was reached. Kinetic modeling of the sequential αTS folding pathway in Fig. 9 *B* was performed with the KINSIM program (55).

Statistical errors reported throughout the article are based on the following grouping of the kinetic trajectories. The 60 trajectories are divided into three groups of 20, i.e., trajectories 1–20, 21–40, 41–60. Properties of each group are averaged and these three separate averages are used to determine a global average. The reported standard deviation shown in Figs. 8 and 9 *A*, and Table 2, is the error of the three group averages.

RESULTS

Simulations of folding thermodynamics

Fig. 1 shows a schematic for reduction of the PDB coordinates 1BKS to the C_α Go-model used for simulations in this study. Fig. 2 shows a representative region of an equilibrium αTS folding/unfolding trajectory this model at 330 K, plotting the number of native contacts, Q, as a function of simulation time steps. From Fig. 2, it is evident that the trajectory significantly samples two intermediate ensembles, I_1 and I_2, between the native ensemble, N, and the unfolded ensemble, U. Intermediates ensembles I_1, corresponding to $Q \sim 280$, and I_2, corresponding to $Q \sim 150$, are sampled in both the unfolding and refolding directions. It should also be noted that the αTS model parameters shown in Table 1 are calibrated to produce a simulation T_f near 335 K, consistent with the experimental T_f, as was shown for chymotrypsin inhibitor 2 in a previous study (24).

TABLE 2 Average folding time (τ) and differences between folding times for kinetic trajectories reaching $Q = 120$, $Q = 200$, $Q = 300$, $Q = 360$, and native $Q = 480$

Average folding time $\langle\tau\rangle$		Folding time difference	
$\langle\tau_{120Q}\rangle$	1.07 ± 0.16	$\langle\tau_{120Q}\rangle - 0$	1.07 ± 0.16
$\langle\tau_{200Q}\rangle$	3.28 ± 1.20	$\langle\tau_{200Q}\rangle - \langle\tau_{120Q}\rangle$	2.21 ± 1.21
$\langle\tau_{300Q}\rangle$	4.82 ± 1.37	$\langle\tau_{300Q}\rangle - \langle\tau_{200Q}\rangle$	1.54 ± 1.82
$\langle\tau_{360Q}\rangle$	6.04 ± 1.26	$\langle\tau_{360Q}\rangle - \langle\tau_{300Q}\rangle$	1.22 ± 1.86
$\langle\tau_{N}\rangle$	7.42 ± 1.92	$\langle\tau_{N}\rangle - \langle\tau_{360Q}\rangle$	1.38 ± 2.30

Errors are means ±SD.

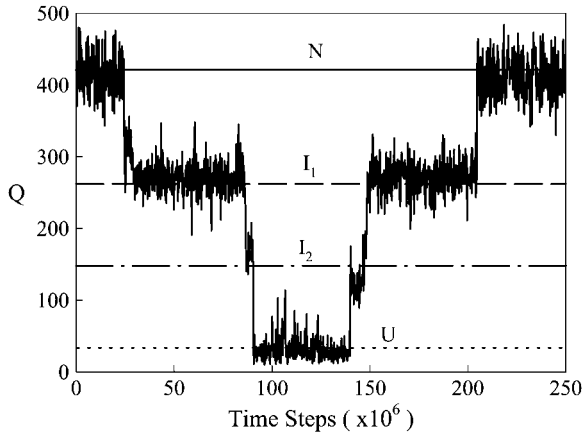


FIGURE 2 The number of native contacts (Q) between 0 and 250×10^6 time steps in an equilibrium simulation of αTS at $T = 330$ K. The simulation samples native N (*solid line* at $Q \sim 430$), intermediate I_1 (*dashed line* at $Q \sim 280$), intermediate I_2 (*dotted-dashed line* at $Q \sim 150$), and unfolded U (*dotted line* at $Q \sim 25$) ensembles.

Through WHAM, the probability of Q values in the equilibrium αTS simulations is used to calculate the PMF of each value of Q sampled during the simulations using Eq. 7 (see Materials and Methods). Fig. 3 shows a plot of *PMF* versus Q determined at the folding temp, $T_f = 335$ K, and shows the relative free-energy differences between states sampled in the simulation. Fig. 3 shows free-energy minima for intermediate ensembles I_1, corresponding to $Q \sim 280$, and I_2, corresponding to $Q \sim 150$. However, Fig. 3 also shows that intermediate ensemble I_2 is actually a broad group of free-energy minima that consists of structures with

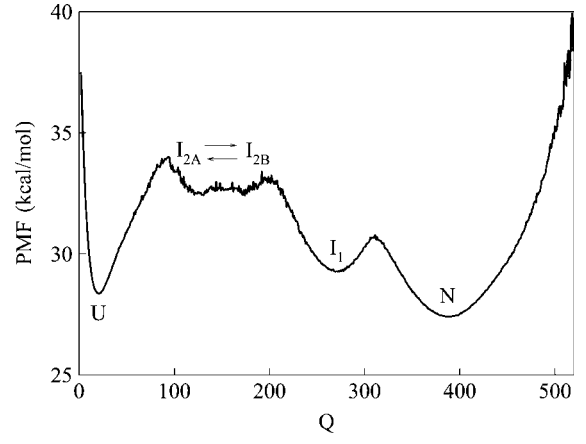


FIGURE 3 A plot of potential mean force (*PMF*) versus native contacts (Q) at the αTS $T_f = 335$ K shows four free-energy minima at the native N ($Q \sim 430$), intermediate I_1 ($Q \sim 280$), intermediate I_{2B} ($Q \sim 180$), intermediate I_{2A} ($Q \sim 120$), and unfolded U ($Q \sim 25$) ensembles. Intermediates I_{2B} and I_{2A} actually occupy the same ensemble intermediate (I_2) and are shown to interconvert (*arrows*).

different Q values. The least structured states are located in a free-energy minimum near $Q \sim 120$, denoted I_{2A} in Fig. 3, and the most structured states are located in a free-energy minimum near $Q \sim 180$, denoted I_{2B} in Fig. 3. It should be noted that error bars of the PMF are not shown due to the lack of repeating unfolding/refolding transitions in the αTS simulations. Although the Q values associated with αTS intermediate ensembles I_1, I_{2A}, and I_{2B}, are easily determined from Figs. 2 and 3, the free-energy differences between these intermediate ensembles remain questionable due to a lack of sampling the folding and unfolding transitions in the equilibrium simulation.

Fig. 4 shows the average value of Q, $\langle Q \rangle$, plotted as a function of simulation temperature. Two unfolding transitions can be clearly identified: 1), N $\rightleftarrows$ I_1 near 310 K and 2), I_1 $\rightleftarrows$ U at 345 K. Due to the low stability of I_{2A} and I_{2B}, transitions involving I_{2A} and I_{2B} are not discernable in Fig. 4. Also indicated in Fig. 4 are the temperature ranges at which each thermodynamic species U, I_{2A} + I_{2B}, I_1, and N, is maximally populated. The simulated titration in Fig. 4 is analogous to chemical or thermal denaturation of αTS in equilibrium experiments.

Fig. 5, *A–D*, shows the structure of the αTS equilibrium intermediate ensembles highlighted in Fig. 4. Fig. 5 *A* shows the fraction of total native contacts formed by each region of secondary structure in equilibrium αTS intermediate ensembles I_{2A}, I_{2B}, and I_1. If an arbitrary definition of 0.2 fraction of native contacts is used to define whether a given region is structured or unstructured, I_{2A} is structured in regions β_1–β_3, I_{2B} is structured in regions α_0–β_4, and I_1 is structured in regions α_0–β_6.

A representative three-dimensional (3D) structure snapshot and a detailed map of native contacts formed in intermediate ensembles I_{2A}, I_{2B}, and I_1 is shown in Fig. 5, *B–D*, respectively. In Fig. 5, *B–D*, squares indicate a native contact as determined from the 1BKS structure with the two residues involved in the contact indicated on the x and y axes. A colored square indicates a native contact, which is formed with >0.5 probability in the intermediate ensemble whereas a black square indicates a native contact formed with <0.5 probability. For clarity, the folded secondary structure regions of I_{2A} (*red*), I_{2B} (*yellow*), and I_1 (*green*) are highlighted in color along the x and y axis. In addition, folded regions of I_{2A} (*red*), I_{2B} (*yellow*), and I_1 (*green*) are highlighted by colored portions of the chain in the shown 3D structure.

In Fig. 5 *B*, the folded β_1–β_3 regions of the intermediate I_{2A} ensemble conformations ($110 < Q < 130$) are indicated by red contact squares in the contact map and highlighted with red chain regions in the representative I_{2A} structure. In Fig. 5 *C*, the folded α_0–β_4 regions of the intermediate I_{2B} ensemble conformations ($170 < Q < 190$) are indicated by yellow contact squares in the contact map and highlighted with yellow chain regions in the representative I_{2B} structure. In Fig. 5 *D*, the folded α_0–β_6 regions of the intermediate I_1 ensemble conformations ($270 < Q < 290$) are indicated by green contact squares in the contact map and highlighted with green chain regions in the representative I_1 structure.

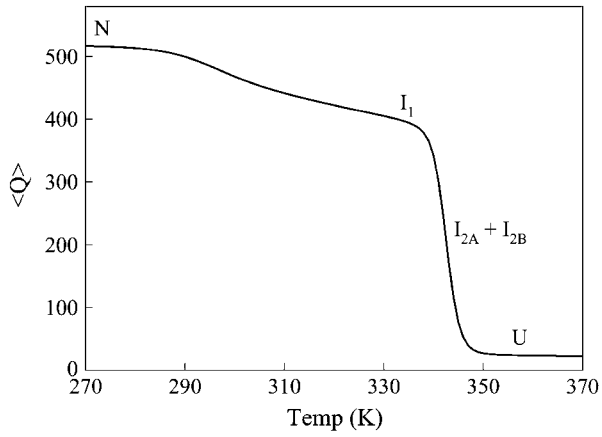


FIGURE 4 The average value of Q, $\langle Q \rangle$, is plotted versus temperature (T) for αTS. Temperatures at which each equilibrium species is maximally populated are shown for native N, intermediate I_1, intermediate I_2 (I_{2A} + I_{2B}), or unfolded U ensembles.

Simulations of refolding kinetics

In addition to equilibrium simulations, 60 kinetic refolding simulation trajectories were run on αTS (see Materials and Methods) to explore the folding of αTS under kinetic conditions. Fig. 6 shows two sample ''native contacts (Q) versus time'' for a fast (*red trace*) and slow (*blue trace*) αTS trajectories. The red trajectory is representative of ~90% of the trajectories and demonstrates progressive folding through a series of transient intermediates. The blue trajectory is representative of ~10% of the trajectories, in which the trajectory is transiently ''trapped'' in an off-pathway intermediate that must unfold completely to refold correctly. With the exception of the off-pathway intermediate state, the intermediate structures sampled by both the red and blue trajectories during refolding appear qualitatively similar. The number of native contacts (Q) corresponding to the native state is indicated by the arrow labeled ''Native'' at $Q = 480$.

To characterize intermediates populated in kinetic refolding simulations of αTS, the probability sampling each value of Q within the first 30 ns (30×10^6 time steps) of all 60 trajectories was calculated and shown in Fig. 7. The y axis of Fig. 7 indicates the probability of adopting a particular conformation with Q native contacts during refolding and is used to identify intermediates populated in refolding pathways. However, this probability cannot be directly related to the actual free energy of the states on these pathways because the probabilities are derived from kinetic, not equilibrium, trajectories. For example, determining the Q probabilities at increasing kinetic simulation lengths at 300 K will naturally increase the probability of sampling native conformations

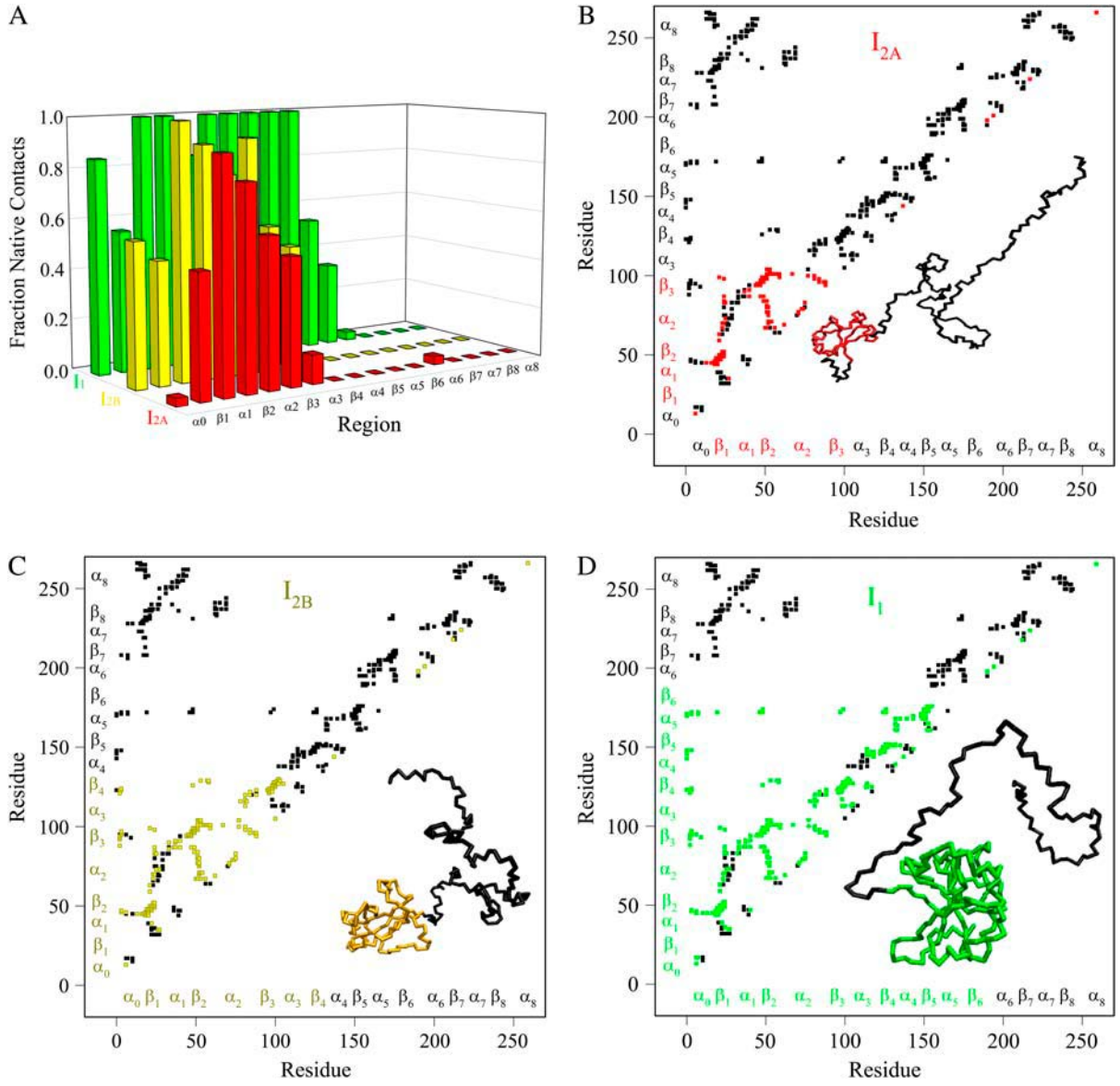


FIGURE 5 Folded regions of αTS equilibrium intermediates. (*A*) Fraction of native contacts formed with probability >0.5 for intermediates I_{2A} (*red bars*), I_{2B} (*yellow bars*), and I_1 (*green bars*) in αTS equilibrium folding and unfolding simulations. (*B*) Contact map and representative MD structure for I_{2A}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{2A} structure with >0.2 fraction of native contacts folded are labeled in red. (*C*) Contact map and representative MD structure for I_{2B}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{2B} structure with >0.2 fraction of native contacts folded are labeled in yellow. (*D*) Contact map and representative MD structure for I_1. Squares in the contact map indicating contacts populated >0.5 and regions in the I_1 structure with >0.2 fraction of native contacts folded are labeled in green. In Fig. 5, *B–D*, squares in the contact maps indicating contacts populated <0.5 and regions of each MD structure with <0.2 fraction of native contacts folded are labeled in black. In Fig. 5, *B–D*, secondary structure elements of αTS are shown along each axis for reference with folded I_{2A} regions colored in red (Fig. 5 *B*), folded I_{2B} regions colored in yellow (Fig. 5 *C*), folded I_1 regions colored in green (Fig. 5 *D*), and unfolded regions colored in black.

over other states, whereas this would not be observed in an equilibrium simulation at 335 K.

From Fig. 7, four distinct kinetic intermediate ensembles are shown to populate significantly between refolding from the unfolded ($Q \sim 25$) to the native ($Q \sim 480$) ensemble in the 60 refolding trajectories of αTS and are located near Q values of 120 (I_{120Q}), 200 (I_{200Q}), 300 (I_{300Q}), and 360 (I_{360Q}). In Fig. 7, designated Q boundaries that define each intermediate ensemble are shown: unfolded U ($10 < Q < 50$, *solid lines*), I_{120Q} ($95 < Q < 145$, *dotted lines*), I_{200Q} ($175 < Q < 225$, *double-dot/dashed lines*), I_{300Q} ($275 < Q < 325$, *single-dot/dashed lines*), I_{360Q} ($335 < Q < 385$, *dashed lines*), and native N ($440 < Q < 520$, *solid lines*). Sampling of these intermediate ensembles can be qualitatively observed in the individual red and blue trajectories shown in Fig. 6.

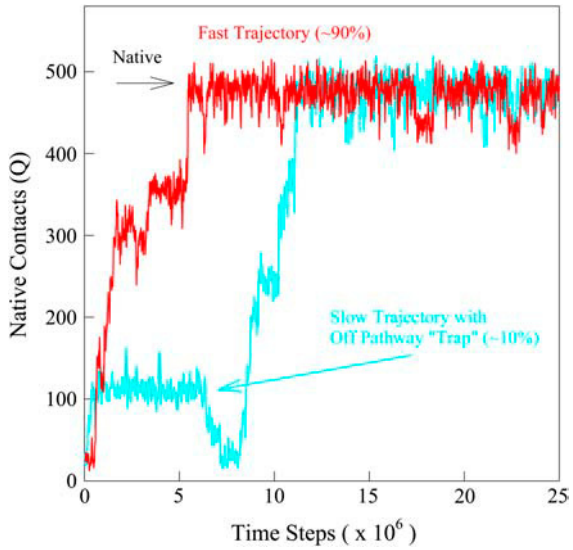


FIGURE 6 The number of native contacts (Q) between 0 and 25 $\times$ 10^6 time steps in a fast-folding (*red*) and slower-folding (*blue*) kinetic refolding trajectory of αTS at $T = 300$ K.

For Fig. 8, *A–E*, the initial 3 $\times$ 10^7 time steps of the kinetic trajectories are segregated into 300 time intervals of 100,000 time steps. In Fig. 8, *A–E*, each of the 60 trajectories is ''counted'' in a time interval bin when it reaches a set value of native contacts (Q) and the total count from all trajectories produces a histogram(56). Fig. 8 *A* shows a histogram of the time intervals in which the 60 trajectories reach $Q = 120$ (approximating I_{120Q}). Fig. 8 *B* shows a histogram of the time intervals in which the 60 trajectories reach $Q = 200$ (approximating I_{200Q}). Fig. 8 *C* shows a histogram of the time intervals in which the 60 trajectories reach $Q = 300$ (approximating I_{300Q}). Fig. 8 *D* shows a histogram of the time intervals in which the 60 trajectories reach $Q = 360$ (approximating I_{360Q}). Fig. 8 *E* shows a histogram of the time intervals in which the 60 trajectories reach $Q = 480$ (approximating Native). To determine the folding time, the average folding time of trajectories 1–20, 21–40, and 41–60 were calculated separately. The average $\langle\tau\rangle$ and standard deviation of these three (1–20, 21–40, 41–60) averaged folding times is shown in Fig. 8, *A–E*, as well as in Table 2 (left column). Table 2 also shows the differences between the folding times between each set of two successive intermediate ensembles (right column). Table 2 indicates that the relative folding time difference between less structured folding ensembles is similar to more structured folding ensembles (right column), although the large relative error of the folding time differences precludes a detailed analysis.

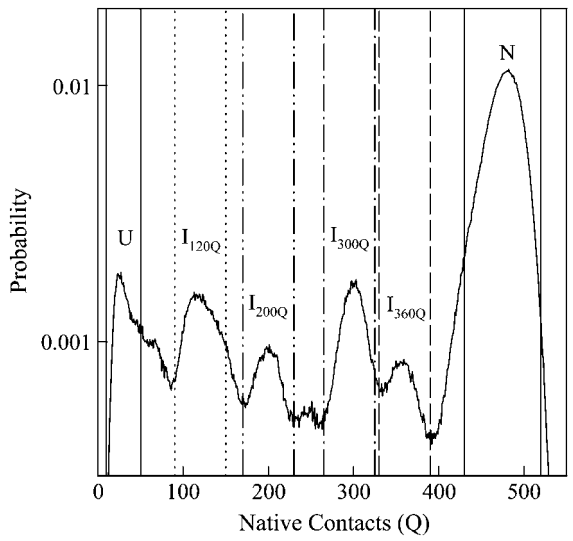


FIGURE 7 The net probability of populating states with Q native contacts during the entirety of the first 30 $\times$ 10^6 time steps of all 60 αTS kinetic refolding trajectories. High probability Q-regions are highlighted between colored lines for unfolded U (*solid lines*), intermediate I_{120Q} (*dotted lines*), intermediate I_{200Q} (*double-dotted/dashed lines*), intermediate I_{300Q} (*single-dotted/dashed lines*), intermediate I_{360Q} (*dashed lines*), and native N (*solid lines*).

A time-resolved description of αTS refolding is shown in Fig. 9 *A*. Using the Q boundary definitions shown in Fig. 7, the probability of adopting each particular ensemble at each simulation time step is determined from all 60 trajectories is indicated in Fig. 9 *A*: unfolded U (*black line*), I_{120Q} (*red line*), I_{200Q} (*yellow line*), I_{300Q} (*green line*), I_{360Q} (*magenta line*), and native N (*black line*). Two important observations of αTS refolding are evident in Fig. 9 *A*. First, the population increase of each folding ensemble is consistently faster for less structured ensembles and a notable ''lag-phase'' is observed for the formation of more structured ensembles (I_{200Q}, I_{300Q}, I_{360Q}, and N). This is consistent with a folding model where the intermediates shown in Fig. 7 are on-pathway and are each formed sequentially throughout the folding process. Second, ~10% of the trajectories appear trapped in the I_{120Q} intermediate ensemble and require more time to refold. A representative trajectory involving one of these I_{120Q} trapped states is shown as a blue trace in Fig. 6.

In Fig. 9 *A*, it should be noted that the fraction of native structures does not equal 1 at 30 $\times$ 10^6 time steps. This observation is due to the fact that three trajectories remain trapped in the I_{120Q} intermediate at this time and require further refolding time to fold completely. In addition, fluctuations in native contacts (Q) in the native ensemble at 300 K occasionally fall outside the range of Q used to define the native ensemble ($440 < Q < 520$). These fluctuations in Q can be observed in Fig. 6 after the red/blue trajectories have reached the native ensemble near $Q \sim 480$.

To provide additional evidence for a sequential folding mechanism in simulations of αTS folding, a pathway involving four sequential intermediates is modeled using KINSIM (55) in Fig. 9 *B* and compared to Fig. 9 *A*. As shown in Fig. 9 *B*, the four kinetic intermediate ensembles observed in Fig. 7 are treated as discrete states in the

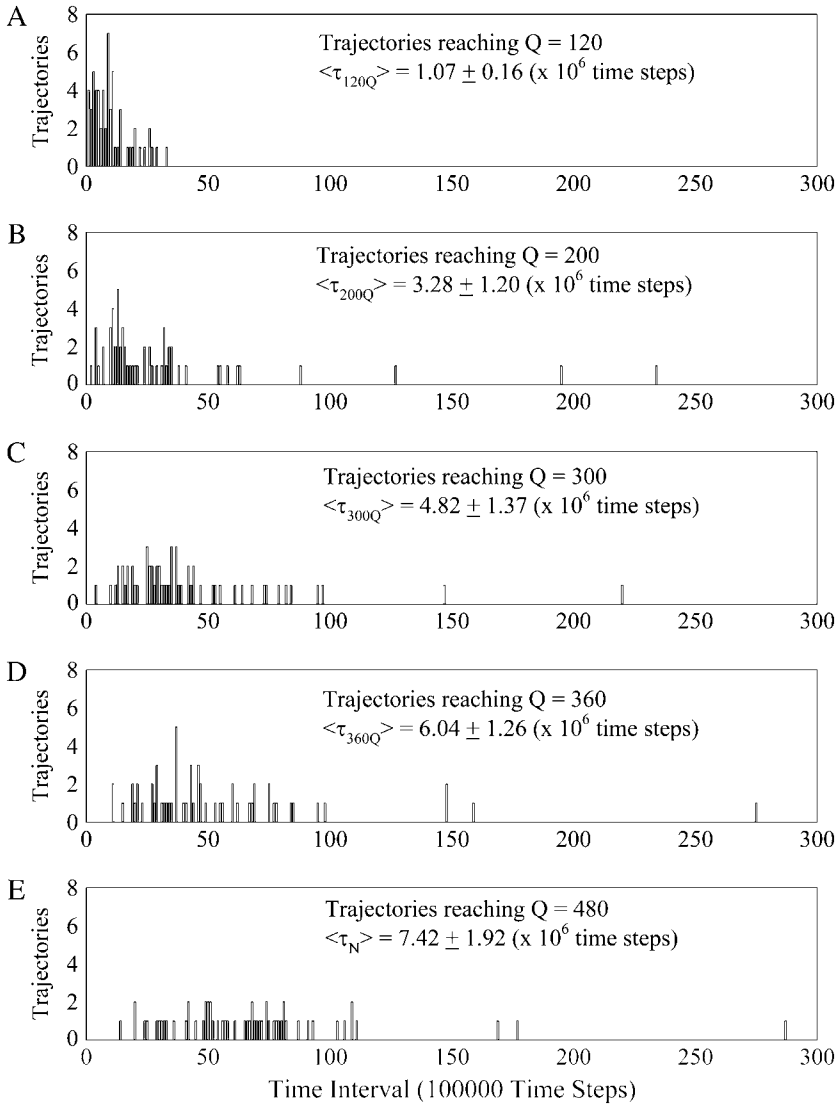


FIGURE 8 The initial 3×10^7 time steps of the kinetic trajectories are segregated into a histogram of 300 time intervals of 100,000 time steps. In Fig. 8, *A–E*, each of the 60 trajectories is ''counted'' in a time interval bin when it reaches $Q = 120$ (*A*), $Q = 200$ (*B*), $Q = 300$ (*C*), $Q = 360$ (*D*), and $Q = 480$ (*E*) native contacts, producing a histogram of folding times. To determine the folding time, the average folding time of trajectories 1–20, 21–40, and 41–60 were calculated separately. The global average, $\langle\tau\rangle$, and standard deviation of the average folding time to reach each Q value is shown in Fig. 8, *A–E* (see Materials and Methods).

following sequential kinetic model that assumes each folding step is irreversible:

$$U \xrightarrow{k_{120Q}} I_{120Q} \xrightarrow{k_{200Q}} I_{200Q} \xrightarrow{k_{300Q}} I_{300Q} \xrightarrow{k_{360Q}} I_{360Q} \xrightarrow{k_N} N.$$

In Fig. 9 *B*, rate constants $k_{120Q} = 3 \times 10^{-7}$, $k_{200Q} = 5 \times 10^{-7}$, $k_{300Q} = 5 \times 10^{-7}$, $k_{360Q} = 3 \times 10^{-7}$, and $k_N = 7 \times 10^{-7}$ produce transient intermediate populations in general agreement with the simulated intermediate populations in Fig. 9 *A*. In particular, the lag phase of intermediate ensemble formation in Fig. 9 *A* is well described using the simple sequential pathway kinetic model in Fig. 9 *B*.

Fig. 10, *A–E*, show the structure of the αTS kinetic intermediate ensembles highlighted in Fig. 7. Fig. 10 *A* shows the fraction of total native contacts formed by each region of secondary structure in kinetic αTS intermediate ensembles I_{120Q}, I_{200Q}, I_{300Q}, and I_{360Q}. If an arbitrary definition of 0.2 fraction of native contacts is used to define whether a given α-helix or β-strand region of αTS is structured or unstructured, I_{120Q} is structured in regions β_1, β_2–β_3, I_{200Q} is structured in regions α_0–β_4, I_{300Q} is structured in regions α_0–β_6, and I_{360Q} is structured in regions α_0–β_7.

A representative 3D structure snapshot and a detailed map of native contacts formed in intermediate ensembles I_{120Q}, I_{200Q}, I_{300Q}, and I_{360Q} is shown in Fig. 10, *B–E*, respectively. In Fig. 10, *B–E*, squares indicate a native contact as determined from the 1BKS structure with the two residues involved in the contact indicated on the *x* and *y* axes. A colored square indicates a native contact, which is formed with >0.5 probability in the intermediate ensemble whereas a black square indicates a native contact formed with <0.5 probability. For clarity, the folded secondary structure regions of I_{120Q} (*red*), I_{200Q} (*yellow*), I_{300Q} (*green*), and

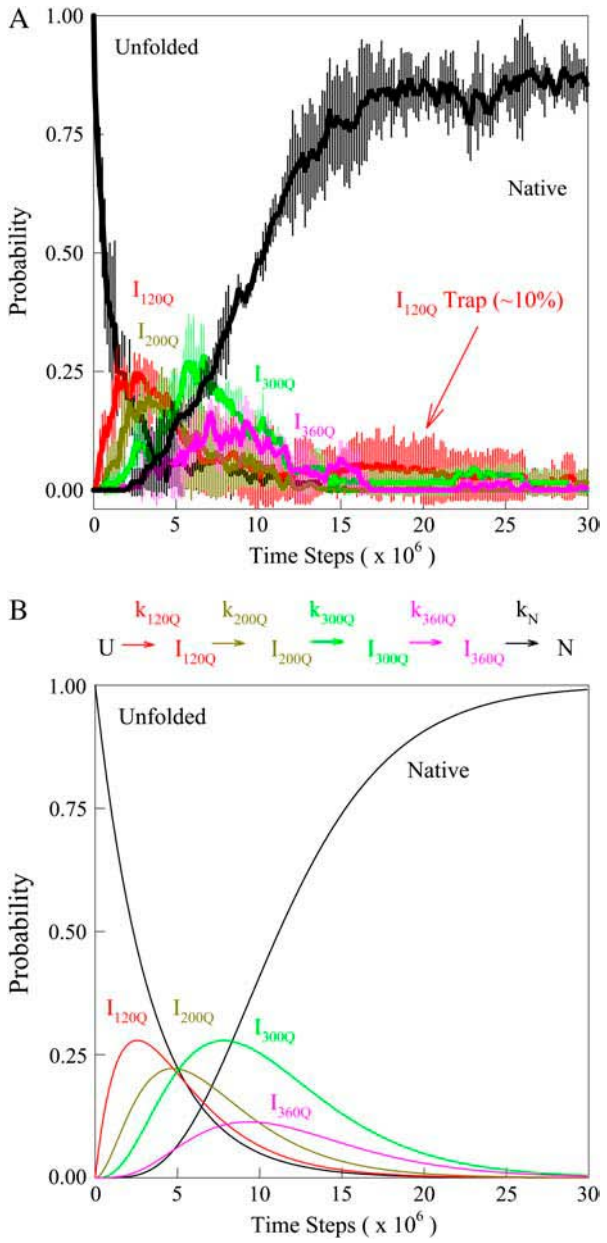


FIGURE 9 (*A*) The probability of populating unfolded, intermediate, and native ensembles at each time point during the simulation. Using the range of Q defined for each ensemble shown in Fig. 7, the probability of populating each ensemble is labeled and color coded for unfolded U (*black*), intermediate I_{120Q} (*red*), intermediate I_{200Q} (*yellow*), intermediate I_{300Q} (*green*), intermediate I_{360Q} (*magenta*), and native N (*black*). A long-lived trapped set of conformations in the I_{120Q} intermediate at long times is indicated with the red arrow labeled ''I_{120Q} trap''. The probability was calculated separately for trajectory groups 1–20, 21–40, and 41–60. The global probability and standard deviation of the three separate probability calculations is shown with error bars at each time point (see Materials and Methods). (*B*) The populations of unfolded, intermediate, and native ensembles shown in Fig. 9 *A* are approximated using a $U \xrightarrow{k_{120Q}} I_{120Q} \xrightarrow{k_{200Q}} I_{200Q} \xrightarrow{k_{300Q}} I_{300Q} \xrightarrow{k_{360Q}} I_{360Q} \xrightarrow{k_N} N$ kinetic model with irreversible rate constants $k_{120Q} = 3 \times 10^{-7}, k_{200Q} = 5 \times 10^{-7}, k_{300Q} = 5 \times 10^{-7}, k_{360Q} = 3 \times 10^{-7}$, and $k_N = 7 \times 10^{-7}$. Probabilities of each species are labeled and color coded as in Fig. 9 *A*, i.e., unfolded U (*black line*), intermediate I_{120Q} (*red*), intermediate I_{200Q} (*yellow*), intermediate I_{300Q} (*green*), intermediate I_{360Q} (*magenta*), and native N (*black*).

I_{360Q} (*magenta*) are highlighted in color along the x and y axis. In addition, folded regions of I_{120Q} (*red*), I_{200Q} (*yellow*), I_{300Q} (*green*), and I_{360Q} (*magenta*) are highlighted by colored portions of the chain in the 3D structure.

In Fig. 10 *B*, the folded β_1, β_2–β_3 regions of the intermediate I_{120Q} ensemble conformations ($95 < Q < 145$) are indicated by red contact squares in the contact map and highlighted with red chain regions in the representative I_{120Q} structure. In Fig. 10 *C*, the folded α_0–β_4 regions of the intermediate I_{200Q} ensemble conformations ($175 < Q < 225$) are indicated by yellow contact squares in the contact map and highlighted with yellow chain regions in the representative I_{200Q} structure. In Fig. 10 *D*, the folded α_0–β_6 regions of the intermediate I_{200Q} ensemble conformations ($275 < Q < 325$) are indicated by green contact squares in the contact map and highlighted with green chain regions in the representative I_{300Q} structure. In Fig. 10 *E*, the folded α_0–β_7 regions of the intermediate I_{360Q} ensemble conformations ($335 < Q < 385$) are indicated by magenta contact squares in the contact map and highlighted with magenta chain regions in the representative I_{360Q} structure.

In Fig. 10 *A*, intermediates of higher Q value (for example, I_{360Q}) demonstrate a greater fraction of contacts formed in each α- or β-region than intermediates of lower Q values (for example, I_{120Q}). In Fig. 10, *B–E*, intermediates of higher Q value (for example, I_{360Q}) always contain contacts of the intermediates with lower Q values (for example, I_{120Q}), in addition to contacts consistent with further structure formation. These observations are highly consistent with a sequential single-pathway folding mechanism.

Simulations and analysis of on- and off-pathway refolding kinetics

Conformations within the intermediate ensemble I_{120Q} appear to be responsible for productive fast folding to the native state (Fig. 6, *red* trajectory) as well as formation of a trapped state that delays proper folding (Fig. 6, *blue* trajectory). Therefore, it is important to compare the structures of the I_{120Q} ensemble found in the 54 fast-folding trajectories with structures of the I_{120Q} ensemble found in the six trapped slow-folding trajectories.

Fig. 11, *A–C*, highlight the structural differences between the fast-folding on-pathway ensemble conformations I^{on}_{120Q} observed in 54 trajectories (*red*) and trapped off-pathway ensemble conformations I^{off}_{120Q} observed in six trajectories (*blue*). Fig. 11 *A* shows the fraction of total native contacts formed by each region of secondary structure in kinetic αTS intermediate ensembles I^{on}_{120Q} and I^{off}_{120Q}. If an arbitrary definition of 0.2 fraction native contacts is used to define whether a given α-helix or β-strand region of αTS is structured or unstructured, I^{off}_{120Q} is structured in regions β_1, β_2–β_3 and I^{off}_{120Q} is structured throughout regions β_1–β_3. Due to its significance to on- and off-pathway folding, the α_2 helix is denoted with a purple asterisk in Fig. 11 *A*.

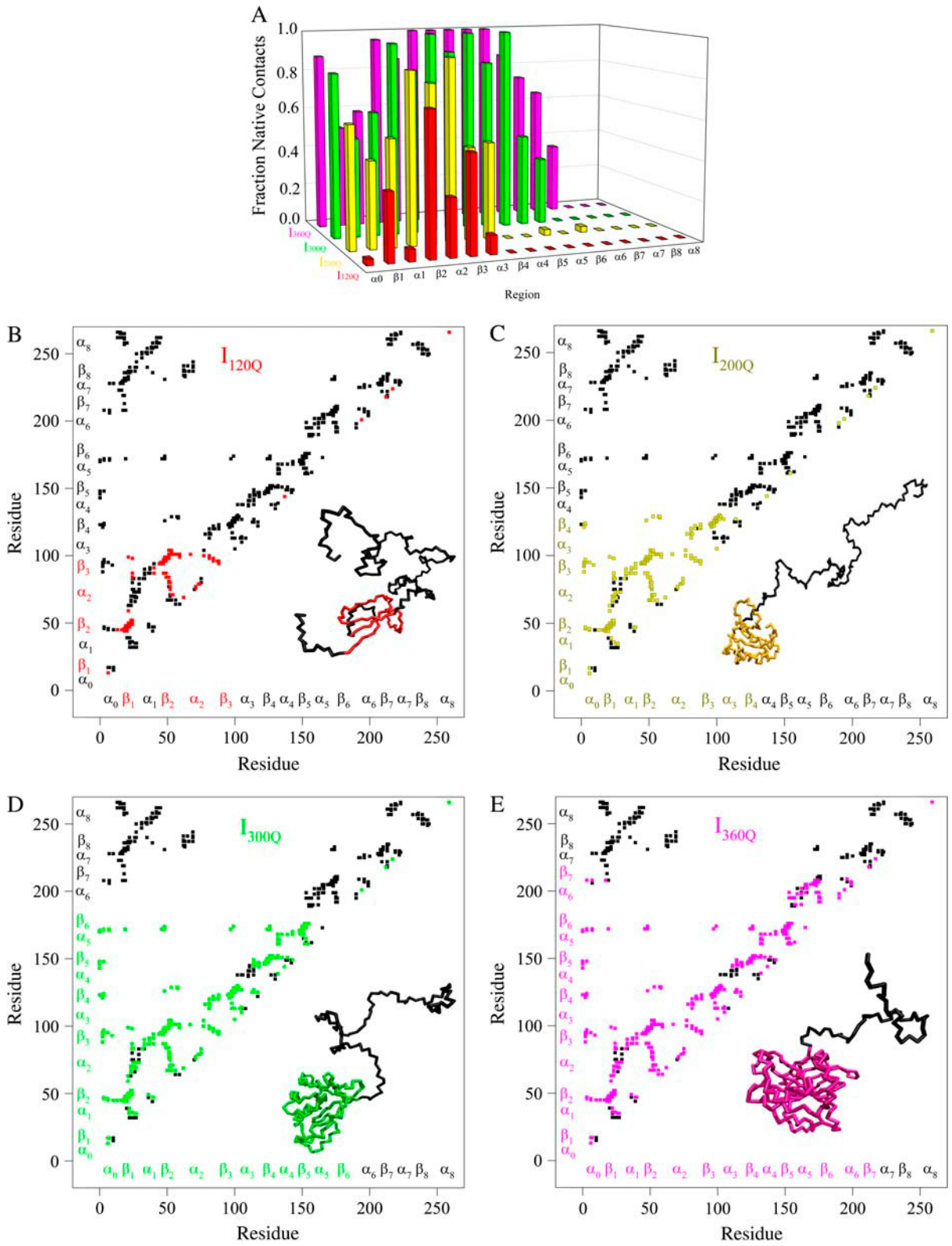


FIGURE 10 Folded regions of αTS kinetic intermediates. (*A*) Fraction of native contacts formed with probability >0.5 for intermediates I_{120Q} (*red bars*), I_{200Q} (*yellow bars*), I_{300Q} (*green bars*), and I_{360Q} (*magenta bars*) in αTS kinetic folding simulations. (*B*) Contact map and representative MD structure for I_{120Q}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{120Q} structure with >0.2 fraction of native contacts folded are labeled in red. (*C*) Contact map and representative MD structure for I_{200Q}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{200Q} structure with >0.2 fraction of native contacts folded are labeled in yellow. (*D*) Contact map and representative MD structure for I_{300Q}. Squares in the contact map

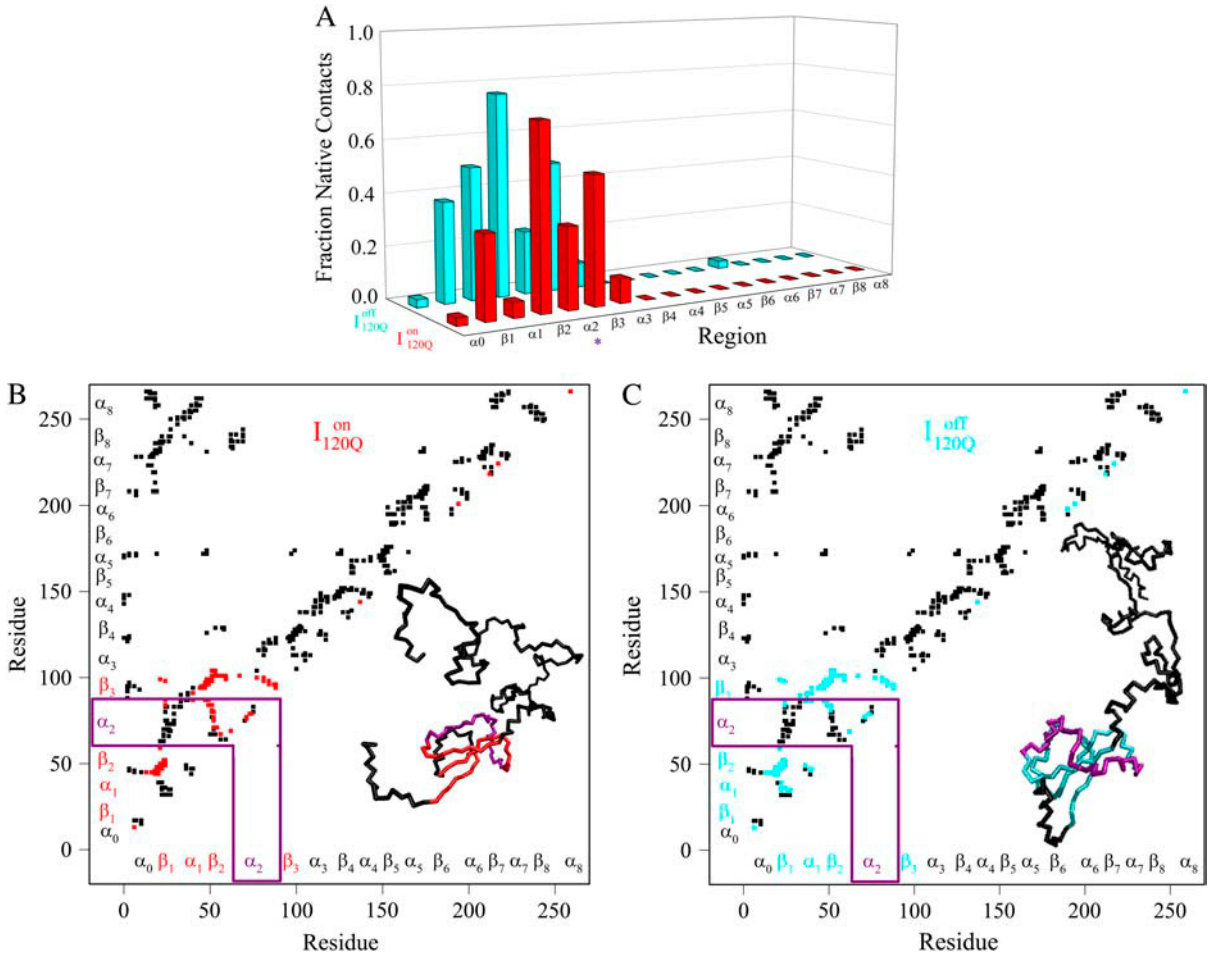


FIGURE 11 Folded regions of αTS kinetic on-pathway intermediate ensemble I^{on}_{120Q} (I_{120Q} from 54 trajectories) and off-pathway intermediate ensemble I^{off}_{120Q} (I_{120Q} from six trajectories). (*A*) Fraction of native contacts formed with probability >0.5 for intermediates I^{on}_{120Q} (*red bars*) and I^{off}_{120Q} (*blue bars*) in αTS kinetic folding simulations. (*B*) Contact map and representative MD structure for I^{on}_{120Q}. Squares in the contact map indicating contacts populated >0.5 and regions in the I^{on}_{120Q} structure with >0.2 fraction of native contacts folded are labeled in red. (*C*) Contact map and representative MD structure for I^{off}_{120Q}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{200Q} structure with >0.2 fraction of native contacts folded are labeled in blue. In Fig. 11, *B* and *C*, squares in the contact maps indicating contacts populated <0.5 and regions of each MD structure with <0.2 fraction of native contacts folded are labeled in black. In Fig. 5, *B* and *C*, secondary structure elements of αTS are shown along each axis for reference with folded I^{on}_{120Q} regions colored in red (Fig. 5 *B*), folded I^{off}_{120Q} regions colored in blue (Fig. 5 *C*), and unfolded regions colored in black.

A representative 3D structure snapshot and a detailed map of native contacts formed in intermediate ensembles I^{on}_{120Q} and I^{off}_{120Q} are shown in Fig. 11, *B* and *C*, respectively. In Fig. 11, *B* and *C*, squares indicate a native contact as determined from the 1BKS structure with the two residues involved in the contact indicated on the *x* and *y* axes. In Fig. 11, *B* and *C*, a colored square indicates a native contact, which is formed with >0.5 probability in the intermediate ensemble whereas a black square indicates a native contact formed with <0.5 probability. For clarity, the folded secondary structure regions of I^{on}_{120Q} (*red*) and I^{off}_{120Q} (*blue*) are highlighted in color along the *x* and *y* axis. In addition, folded regions of I^{on}_{120Q} (*red*) and I^{off}_{120Q} (*blue*) are highlighted by colored portions of the chain in the 3D structure. Due to the significance of the

FIGURE 10 (Continued).
indicating contacts populated >0.5 and regions in the I_{300Q} structure with >0.2 fraction of native contacts folded are labeled in green. (*E*) Contact map and representative MD structure for I_{360Q}. Squares in the contact map indicating contacts populated >0.5 and regions in the I_{360Q} structure with >0.2 fraction of native contacts folded are labeled in magenta. In Fig. 5, *B–E*, squares in the contact maps indicating contacts populated <0.5 and regions of each MD structure with <0.2 fraction of native contacts folded are labeled in black. In Fig. 5, *B–E*, secondary structure elements of αTS are shown along each axis for reference with folded I_{120Q} regions colored in red (Fig. 5 *B*), folded I_{200Q} regions colored in yellow (Fig. 5 *C*), folded I_{300Q} regions colored in green (Fig. 5 *D*), folded I_{360Q} regions colored in magenta (Fig. 5 *E*), and unfolded regions colored in black.

α_2 helix in modulating on- and off-pathway folding, contacts between the α_2 helix and other folded regions of I^{on}_{120Q} and I^{off}_{120Q} are located within a purple boundary in the contact map and the portion of the 3D structure corresponding to the α_2 helix is also colored purple.

In Fig. 11 *B*, the folded β_1, β_2–β_3 regions of the intermediate I^{on}_{120Q} ensemble conformations ($95 < Q < 145$ from 54 on-pathway trajectories) are indicated by red contact squares in the contact map and highlighted with red chain regions in the representative I^{on}_{120Q} structure. In Fig. 11 *C*, the folded β_1–β_3 regions of the intermediate I^{off}_{120Q} ensemble conformations ($95 < Q < 145$ from six off-pathway trajectories) are indicated by blue contact squares in the contact map and highlighted with blue chain regions in the representative I^{off}_{120Q} structure.

In Fig. 11 *A*, it is interesting to note that contacts in α-helix α_1 are nearly absent in I^{on}_{120Q} but largely present in I^{off}_{120Q}. In contrast, contacts in α-helix α_2 (indicated by *purple asterisk*) are slightly higher in I^{on}_{120Q} than in I^{off}_{120Q}. In Fig. 11, *B* and *C*, it is observed which contacts explain these structural differences exist between I^{on}_{120Q} and I^{off}_{120Q}. Comparing Fig. 11, *B* and *C*, one can observe that I^{on}_{120Q} lacks contacts formed between α_1 and α_2 and that I^{on}_{120Q} lacks contacts formed between the C-terminus of β_2 and the N-terminus of α_2. Comparing the structures of I^{on}_{120Q} and I^{off}_{120Q} in Fig. 11, *B* and *C*, it can be clearly observed that helix α_2, colored in purple, incorrectly wraps clockwise around strand β_3 in I^{off}_{120Q}. This nonnative chiral arrangement of α_2 is stabilized by native contacts formed between the N-terminus of β_2 and the C-terminus of α_2. This nonnative I^{off}_{120Q} conformation also places α_2 on the face of β_3 where the next strand, β_4, would dock in the subsequent intermediate ensemble I_{200Q}. The location of the α_2 helix in the I^{off}_{120Q} conformation would sterically prevent stable assembly of β_3 and β_4, resulting in a dead-end folding intermediate.

If the I^{off}_{120Q} conformation is a dead-end conformation, unfolding is required to correctly refold. To test whether I^{off}_{120Q} conformations are dead ends, refolding kinetic simulations were conducted initiating with either an I^{on}_{120Q} and I^{off}_{120Q} conformation. For I^{on}_{120Q} conformations, five conformations with $Q = 120$ were randomly selected from six on-pathway trajectories (trajectories 6, 34, 45, 47, 57, and 59) as initial structures for refolding. For I^{off}_{120Q} conformations, five conformations with $Q = 120$ were randomly selected from the six off-pathway trajectories (trajectories 7, 35, 46, 48, 58, and 60) as initial structures for refolding. These structures are given random initial velocities and refolded at 300 K.

Fig. 12 shows the results of the 30 simulations initiating with either I^{on}_{120Q} (*red symbols*) or I^{off}_{120Q} (*blue symbols*) conformations. For each trajectory, Fig. 12 plots the minimum value of Q sampled on the x axis versus the folding time, the time to reach the native ensemble $Q = 480$, on the y axis. Symbol shapes indicate initial conformations obtained from trajectories 6,7 (*circles*), 34,35 (*squares*), 45,46 (*triangles*), 47,48 (*inverted triangles*), 57,58 (*diamonds*), and 59,60 (*hexagons*). If I^{off}_{120Q} is truly a dead-end structure, trajectories initiating from I^{off}_{120Q} should sample values of Q much lower than I^{on}_{120Q} before refolding, because they would have to unfold before correct refolding. Because unfolding is required before refolding, trajectories initiating from I^{off}_{120Q} would also be expected to take longer to properly refold.

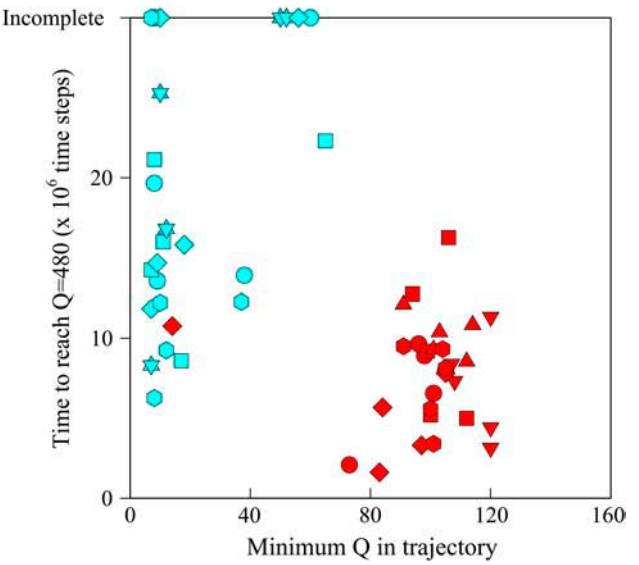


FIGURE 12 Time steps to reach native ensemble ($Q = 480$) and minimum value of Q sampled during simulations initiating with $Q = 120$ structures from I^{on}_{120Q} (*red symbols*) and I^{off}_{120Q} (*blue symbols*) ensembles. Five different initial I^{on}_{120Q} configurations were obtained from trajectories 6 (*circles*), 34 (*squares*), 45 (*triangles*), 47 (*inverted triangles*), 57 (*diamonds*), and 59 (*hexagons*). Five different initial I^{off}_{120Q} configurations were obtained from trajectories 7 (*circles*), 35 (*squares*), 46 (*triangles*), 48 (*inverted triangles*), 58 (*diamonds*), and 60 (*hexagons*).

In Fig. 12, trajectories initiating from I^{on}_{120Q} and I^{off}_{120Q} clearly show significant differences and cluster in different regions of the plot. As expected, all I^{off}_{120Q} trajectories sample lower values of Q ($Q < 70$) than all I^{on}_{120Q} trajectories ($Q > 70$), with one exception of an I^{on}_{120Q} trajectory that appears to unfold. This observation clearly shows that I^{off}_{120Q} conformations require a significant degree of unfolding before refolding. Although some I^{off}_{120Q} trajectories refold to the native ensemble at times comparable to I^{on}_{120Q} trajectories, the average refolding time is clearly slower for I^{off}_{120Q} trajectories. The fastest trajectories initiate from I^{on}_{120Q} conformations whereas the 14 slowest trajectories initiate from I^{off}_{120Q} conformations, with nine of these trajectories remaining unfolded at the end of the 30×10^6 time steps (labeled ''incomplete'' on the y axis). There do not appear to be significant differences between values of different symbols of each I^{on}_{120Q} (*red*) and I^{off}_{120Q}(*blue*) trajectory subgroups, indicating that each structural subgroup is largely homogeneous with respect to folding.

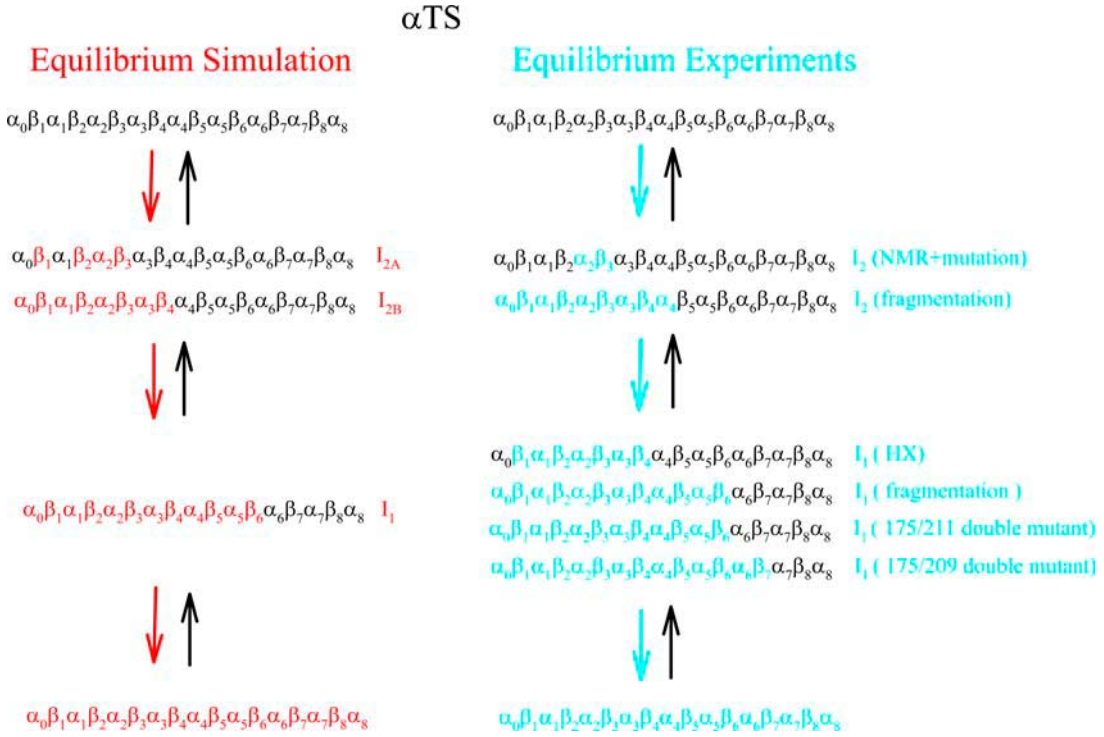
DISCUSSION

Equilibrium simulations of αTS agree with experiments

Figs. 2 and 3 indicate that equilibrium folding/unfolding of αTS at $T_f = 335$ K occurs with the population of two predominant intermediate ensembles, I_1 and I_2, in an apparently sequential process. Fig. 5, *A–C*, shows that structures within the intermediate I_2 ensemble (I_{2A}–I_{2B}) show contacts in the N-terminal region, although the I_2 structures occupy a broader distribution of Q values than I_1. The least structured of the I_2 ensemble (I_{2A}) shows native contacts in the β_1–β_3 regions whereas the most structured (I_{2B}) shows native contacts throughout the α_0–β_4 regions. Fig. 5, *A* and *D*, also shows that I_1 consists of a discrete ensemble of intermediate structures with native contacts in the N-terminal α_0–β_6 regions and the C-terminal region α_6–α_8 unstructured. Fig. 5, *B–D*, show that, within the folded region of each equilibrium intermediate ensemble (i.e., *colored squares*), both short-range and longer-range contacts are equally probable. Thus, even though unfolded regions still exist in these intermediates, the secondary (short-range) and tertiary (long-range) structure within the folded region of the αTS chain fold concomitantly. In Scheme 1, these results are compared to equilibrium experimental measurements on αTS.

Scheme 1 shows that αTS equilibrium folding simulations agree with αTS equilibrium folding experiments. The simulations produce two dominant free-energy minima, I_1 and $I_{2A} + I_{2B}$, in equilibrium folding, which agree well with the two intermediates identified from fits of experimental urea titrations (39). The simulation structure of intermediate I_1 (α_0–β_6) is highly similar to the structure of I_1 inferred from fragment stability experiments (α_0–β_6) (40), a Y175C/G211E double mutant perturbation study (α_0–β_6) (51), a Y175Q/L209V double mutant perturbation study (α_0–β_7) (45), although it is larger than I_1 structural regions exhibiting HX protection (β_1–β_4) (37).

In Scheme 1, conformations contributing to the free-energy minima between I_{2A} and I_{2B} also agree well with experimental studies probing the structure of intermediate I_2 (38,40). The earliest regions of folding in the simulation, denoted as intermediate I_{2A}, predict initial folding in the N-terminus (β_1–β_3), consistent with structured regions in I_2 measured with NMR and probed with changes in stability upon single-site mutagenesis (α_2, β_3) (38). The most structured of states in the I_2 intermediate ensemble in the simulations, denoted I_{2B}, indicate continuous structure in regions α_0–β_4. This finding is also consistent with fragmentation experiments that show that αTS fragments shorter than α_0–α_4 no longer fold (40). In addition, although HX experiments have not been able to detect HX protection under conditions where the I_2 intermediate is maximally populated, they do indicate an HX-resistant region of I_1 in regions β_1–β_4 that may reflect a stable core reflective of the I_2 structure (37). The consistency between the funneled energy model with these experiments supports the



SCHEME 1 Comparison of equilibrium folding simulations and experiments for αTS.

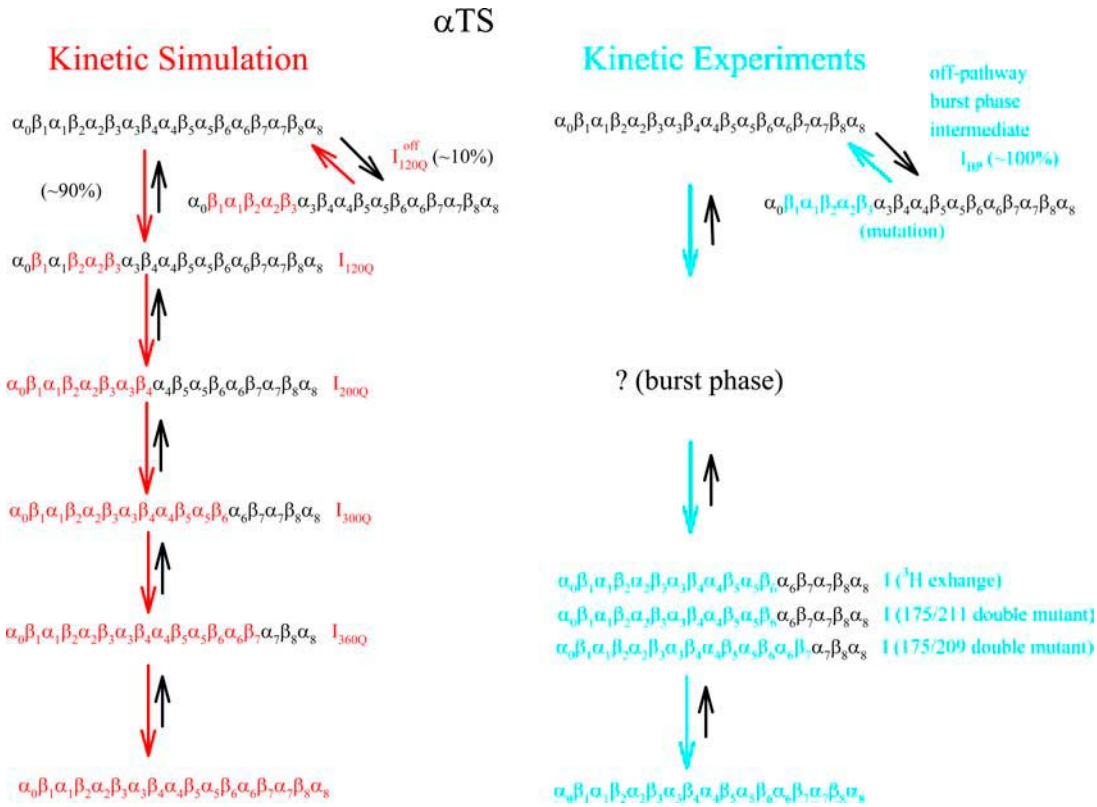
hypothesis that the energy landscape of αTS is highly funneled to the native state.

Kinetic simulations of αTS produce a sequential folding mechanism and folding intermediates consistent with experiments

Figs. 6, 7, 8, *A–E*, and 9 *A* indicate that kinetic folding of αTS at 300 K ($\sim$0.9 T_f) occurs with the population of four predominant on-pathway intermediate ensembles (I_{120Q}, I_{200Q}, I_{300Q}, and I_{360Q}) and one off-pathway intermediate (denoted as I^{off}_{120Q}) in a sequential process. Fig. 10 *B* shows that on-pathway I_{120Q} consists of initial structure formation in the N-terminal β_1, β_2–β_3 regions. Fig. 10 *C* shows that the subsequent intermediate, I_{200Q}, contains the same contacts as I_{120Q}, with additional contacts between regions α_0–β_4. Fig. 10 *D* shows that the next intermediate, I_{300Q}, contains the same contacts as I_{200Q}, with additional contacts between regions α_0–β_6. Fig. 10 *E* shows that the final intermediate, I_{360Q}, contains the same contacts as I_{300Q}, and with additional contacts between regions α_0–β_7.

For the αTS equilibrium simulations, an on-pathway sequential folding mechanism appears to be the case because both folding and unfolding events sample similar intermediate *Q* ensembles (Fig. 2). A sequential on-pathway folding mechanism is also consistent with 90% of αTS kinetic simulations and is supported by three observations. First, visual inspection of individual trajectories indicates sequential formation of the four intermediate ensembles during refolding (*red* trajectory shown in Fig. 6). Second, the increasing lag phase shown, respectively, for αTS intermediates I_{200Q}, I_{300Q}, I_{360Q}, and native state N in Fig. 9 *A* is consistent with the sequential pathway model shown in Fig. 9 *B*. Third, each successively more structured intermediate ensemble in Fig. 10, *A–E*, contains all the contacts of the previous intermediate ensemble. If these intermediates were each populated in a parallel pathway, this property might not be observed.

Scheme 2 shows that αTS kinetic folding simulations agree with the results of the αTS kinetic folding experiments conducted up to this point. Before comparison, it is important to note that the parallel pathways resulting from proline isomerization are not observed in the αTS kinetic refolding simulations (41,42). Although it is possible that proline isomerization could have been captured with this simple topological model of αTS, it is not surprising that isomerization was not observed in the simulations because proline isomerization was not specifically engineered into the model. Due to the complications of proline isomerization and a significant stopped flow signal within the dead time of mixing, a discrete number and structures of early αTS folding intermediates is difficult to determine (41,42).



SCHEME 2 Comparison of kinetic folding simulations and experiments for αTS.

However, kinetic experiments of αTS suggest that at least one intermediate, with a structural ensemble similar to I_1 in equilibrium studies (α_0–β_7) is populated during kinetic refolding (30,41,42,45,51).

Determining whether intermediates I_{120Q}, I_{200Q}, and I_{300Q}, populated in kinetic simulations, are also populated in kinetic experiments is difficult because early αTS refolding intermediates populate in the dead time of kinetic mixing. However, the structure of the final intermediate in simulations, I_{360Q} (α_0–β_7), can be compared to the final intermediate detected in experiments before native state formation. An early ^{3}H exchange experiment revealed that this intermediate is significantly more protected in the N-terminal regions α_0–β_6 than the C-terminal regions α_6–α_8 (30). Furthermore, double mutant studies suggest that contacts between Y175 (β_6) and L209 (β_7) partially exist in the kinetic intermediate preceding native state formation (45), although contacts between Y175 (β_6) and G211 (β_7) were not detected for this intermediate in a similar study (51). The excellent structural agreement is found between simulation intermediate I_{360Q} and the structure inferred from kinetic experiments (α_0–β_7) is shown in Scheme 2.

The relative folding rates between on-pathway unfolded, intermediate, and native ensembles appear to be qualitatively similar throughout all the simulated kinetic trajectories. Table 2 shows quantitatively that differences between the average folding time do not vary significantly between different intermediate ensembles. For example, the folding time difference between the early intermediate ensembles, $\langle\tau_{300Q}\rangle - \langle\tau_{200Q}\rangle = 1.54 \times 10^6$, is similar to the final two species, $\langle\tau_N\rangle - \langle\tau_{360Q}\rangle = 1.38 \times 10^6$. This observation indicates that transition times throughout simulated αTS folding are similar. Furthermore, the simple kinetic model shown in Fig. 9 *B* produces similar intermediate populations during αTS refolding as observed from MD simulations in Fig. 9 *A* with similar rate constants for each kinetic step ($k = 3$–7×10^{-7} time steps^{-1}).

It should be noted that similar rates between different simulated αTS folding steps is not entirely consistent with different folding rates fitted in kinetic experimental data ($k_{UI} \sim 100$ s^{-1} and $k_{IN} \sim 10$ s^{-1}) (41). One explanation for this discrepancy may be that the minimalist C_α model may not accurately capture the side-chain packing contribution to kinetic barrier heights at increasingly structured intermediate stages of αTS folding. Also, future mutational studies are needed to eliminate the off-pathway and proline isomerization steps to facilitate a direct rate measurement of k_{UI} and k_{IN} (42). Regardless of these rate discrepancies between simulations and experiments, the fact that the simulated intermediate populations (Fig. 9 *A*) are sufficiently modeled with a simple sequential kinetic model (Fig. 9 *B*) provides evidence for a predominantly sequential folding of αTS in the kinetic trajectories. The sequential nature of the on-pathway folding mechanism agrees qualitatively with experiments, neglecting proline isomerization (41).

Kinetic simulations capture the off-pathway intermediate observed in kinetic refolding experiments

Off-pathway intermediates have been shown in a number of theoretical (57), computational (8), and experimental (58,59) studies of protein folding. In αTS simulations, 10% of trajectories initially fold to a trapped intermediate, I^{off}_{120Q}, which requires a relatively slow unfolding process to occur to properly refold through the productive sequential folding channel (*blue* trajectory shown in Fig. 6). Evidence for this trapped ensemble I^{off}_{120Q} is demonstrated by the persistent presence of 0.1 fraction of the I_{120Q} intermediate in Fig. 9 *A* and the resistance of this species to further folding. Although Fig. 11 *A* shows that the structures of on-pathway I^{on}_{120Q} and the off-pathway I^{off}_{120Q} both show contacts between β-strands β_1–β_3, there are very few helix α_1 contacts in I^{on}_{120Q} and slightly fewer contacts between strand β_2 and helix α_2 in I^{off}_{120Q}. A comparison of the structures of I^{on}_{120Q} (Fig. 11 *B*) and I^{off}_{120Q} (Fig. 11 *C*) reveals that, helix α_2, colored in purple, is wrapped in a nonnative clockwise chiral arrangement around β_3 in I^{off}_{120Q}. Steric repulsion between α_2 and the next docking unit $\alpha_3\beta_4$ prevents further folding unless an unfolding event would allow α_2 to disassemble and refold in the correct orientation. Fig. 12 confirms that I^{off}_{120Q} conformations must at least partially unfold to correctly refold.

It is striking that the kinetic simulations in this study identify this off-pathway intermediate I^{off}_{120Q} because an off-pathway intermediate I_{BP} is also observed in kinetic experiments (41). Scheme 2 shows that the mechanism by which I^{off}_{120Q} is folded and slowly unfolded in simulations is similar to that proposed based on experimental kinetic studies of I_{BP} (41). Although kinetic simulations indicate that only a minor fraction ($\sim$0.1) of the trajectories sample the off-pathway state, I^{off}_{120Q}, experiments suggest that nearly all molecules should sample this off-pathway intermediate. Nonetheless, it is remarkable that the Go-model of αTS captures this off-pathway intermediate trap because the attractive contacts in the Go-model are determined solely from native contacts observed in the PDB structure 1BKS. Furthermore, perturbation of the stability of the off-pathway intermediate I_{BP} with single-site mutations provides evidence for structure of I_{BP} in regions β_1–β_3 (C. R. Matthews, University of Massachusetts Medical School, personal communication). This result is in excellent agreement with these simulations of the αTS Go-model, which demonstrate that the off-pathway kinetic intermediate, I^{off}_{120Q}, also makes contacts between regions β_1–β_3 (Fig. 11 *C*).

The nature of the off-pathway trap I^{off}_{120Q} is a nonnative chiral topology in the N-terminal bend of helix α_2 held in place by native contacts between the ''ends'' of the structural unit (N-terminus of β_2 and C-terminus of α_2) (Fig. 11 *C*).

This topological arrangement prevents further assembly of αTS folding units beyond strand β_3 and requires breaking of at least 50 contacts to correctly refold (*blue symbols* in Fig. 12). Evidence for this structural hypothesis for I^{off}_{120Q} formation involving the nonnative configuration of helix α_2 shown in Fig. 11 *C* may be determined in future mutational studies of αTS folding kinetics. In addition, future computational work will be aimed at developing an order parameter that is able to discern whether an intermediate conformation similar to I_{120Q} is folding competent (I^{on}_{120Q}) or less ''foldable'' than most unfolded conformations (I^{off}_{120Q}). For this study, a complete equilibrium and kinetic analysis proved sufficient to identify the off-pathway intermediate conformation I^{off}_{120Q} using the Q order parameter.

The αTS folding pathway is robust between equilibrium and kinetic simulations

Comparison of the simulated αTS equilibrium folding pathway in Scheme 1 and simulated αTS kinetic folding pathway in Scheme 2 shows that αTS equilibrium and kinetic simulation pathways are highly similar. Despite the observation that the kinetic pathway appears to form four intermediate ensembles with an off-pathway intermediate and the equilibrium ensemble appears to form only two free-energy minima and no off-pathway intermediates, there are significant structural similarities between intermediates populated in these two pathways. Comparison of equilibrium and kinetic intermediates is based upon the whether α-helices and β-sheets are at least partially folded, as defined by whether a given α-helix or β-sheet has >0.2 fraction of native contacts formed with >0.5 probability (Figs. 5 *A* and 10 *A*).

First, the equilibrium intermediate I_{2A} (Fig. 5, *A* and *B*) shows folding throughout regions β_1–β_3, which is highly similar to kinetic intermediate I_{120Q} (Fig. 10, *A* and *B*), which is folded in four of these five structural regions (β_1, β_2, α_2, β_3). Second, the equilibrium intermediate I_{2B} (Fig. 5, *A* and *C*) shows folding throughout regions α_0–β_4, which is identical to folded regions of kinetic intermediate I_{200Q} (Fig. 10, *A* and *C*). Third, the equilibrium intermediate I_1 (Fig. 5, *A* and *D*) shows folding throughout regions α_0–β_6, which is identical to folded regions of kinetic intermediate I_{300Q} (Fig. 10, *A* and *D*). Although no equilibrium intermediate was identified that matched the structure of the kinetic intermediate I_{360Q}, I_{360Q} is the least-populated kinetic ensemble in Fig. 7 ($T = 300$ K) and may not be stable during equilibrium simulations at higher temperature ($T = 335$ K). Interestingly, I_{360Q} identifies β7 contacts consistent with those detected in both kinetic and equilibrium experiments using a double-mutant thermodynamic cycle (45). This agreement emphasizes the importance of using both kinetic and equilibrium simulations to explore protein folding pathways. In any case, the energy landscape of αTS appears to be extremely robust, such that the folding pathway is only minimally altered when αTS refolds and unfolds in reversible equilibrium at 335 K (T_f) (Fig. 2) or refolds irreversibly in kinetics at 300 K (0.9 T_f) (Fig. 6).

Comparison to previous TIM barrel simulations

It is also important to compare simulated αTS kinetics in this study with those in a previous kinetic simulation study of αTS (23). Certain energetic and entropic differences exist between the minimalist C_α Go-model in this study, in which the native state is defined as the lowest energy state, and the all-atom model of Godzik et al., which energetically biases local backbone atoms to the native dihedrals but permits nonnative long-range contacts with a Miyazawa-Jernigan weighted statistical potential (23). Also, this study uses off-lattice molecular dynamics whereas the previous all-atom study uses an on-lattice Monte-Carlo method.

Despite differences in the energy functions used, the results of this Go-model and the all-atom model are surprisingly consistent. First, both studies predict significant population of N-terminal kinetic intermediate ensembles, corresponding to I_{300Q} in this study (Fig. 10), which are also found in experiments (30,41,45,51). Second, both models appear to model the trapped off-pathway structures observed in experiments (41).

However, a few important differences do exist between this study and the all-atom study. First, this model predicts the earliest kinetic folding events in a region corresponding to residues 15–110 (intermediate ensemble I_{120Q}) whereas the all-atom study shows much of this area folding later than the neighboring contacts (23). Since equilibrium analysis has shown the I_{120Q} ''contact cluster'' to be metastable (intermediate I_{2A} and I_{2B} in Fig. 3), small differences in the on-lattice model may underweight the presence of the I_{120Q} ensemble. It is difficult to confirm the presence and structure of the putative I_{120Q} on-pathway kinetic intermediate with experiments because these intermediates would be expected to populate in the burst phase of stopped flow mixing and also copopulate with the unfolded ensemble (41).

Second, although both studies appear to capture off-pathway trapped states, significant differences exist as to the stability and structure of these off-pathway traps. In this study, the off-pathway intermediate, I^{off}_{120Q}, is observed in 10% of the simulations and corresponds to N-terminal regions β_1–β_3. In the all-atom study, nearly half of the simulations appear trapped at the end of the run (native *RMS* > 8 Å) and appear to involve trapped structures involving both N- and C-terminal regions (50% N-terminal trap/50% C-terminal trap) (23). Recent mutational experiments support a trapped structure involving contacts between N-terminal regions β_1–β_3 and do not appear to involve C-terminal contacts (C. R. Matthews, University of Massachusetts Medical School, personal communication).

Regardless of these discrepancies, it is not surprising that differences are observed between the two models. These differences might be attributed to the fact that the previous lattice model uses explicit atoms and nonnative contacts whereas this C_α Go-model does not. Despite these differences, the intriguing similarity between the two models of αTS may help explain the ubiquitous occurrence of the TIM barrel fold across genomes. If the native fold and folding pathway, determined in this study using a native Go-model, can also be determined from simulations without a priori knowledge of native contacts (23), it is possible that the evolutionary selection process of protein structure also finds the TIM barrel fold a readily accessible motif. In terms of landscape theory, the TIM barrel appears to be a relatively uncomplicated fold with a low chain entropy penalty paid during folding. Alternatively, this fold could have evolved due to its functional versatility in the catalysis of metabolic reactions. Further studies of other members of the TIM barrel fold class will be necessary to address the degree to which small changes in the native topology of different TIM barrel proteins impact the folding of simple computational models, as well as real proteins.

We thank Dr. C. Robert Matthews, Dr. Ramakrishna Vadrevu, Dr. Ying Wu, and Dr. Osman Bilsel at the University of Massachusetts Medical School, Dept. of Biochemistry and Molecular Pharmacology, for helpful discussion and critical review of the manuscript. Additional computational support has been provided by the W. M. Keck Foundation and the Keck II Center at University of California at San Diego.

We acknowledge financial support from National Science Foundation grants MCB-0084797, PHY-0216576, and PHY-0225630, and the National Institutes of Health Postdoctoral Fellowship GM064936-01 (J.M.F.).

REFERENCES

1. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*. 84:7524–7528.
2. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
3. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*. 89:8721–8725.
4. Garel, T., and H. Orland. 1988. Mean-field model for protein folding. *Europhys. Lett.* 6:307–310.
5. Shakhnovich, E. I., and A. M. Gutin. 1989. The nonergodic (spin-glass-like) phase of heteropolymer with quenched disordered sequence of links. *Europhys. Lett.* 8:327–332.
6. Dill, K. A., S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. 1995. Principles of protein folding—a perspective from simple exact models. *Protein Sci.* 4:561–602.
7. Karplus, M., and A. Sali. 1995. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73.
8. Camacho, C. J., and D. Thirumalai. 1996. Denaturants can accelerate folding rates in a class of globular proteins. *Protein Sci.* 5: 1826–1832.
9. Marqusee, S., V. H. Robbins, and R. L. Baldwin. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA*. 86:5286–5290.
10. Munoz, V., P. A. Thompson, J. Hofrichter, and W. A. Eaton. 1997. Folding dynamics and mechanism of beta-hairpin formation. *Nature*. 390:196–199.
11. Yang, W. Y., J. W. Pitera, W. C. Swope, and M. Gruebele. 2004. Heterogeneous folding of the trpzip hairpin: full atom simulation and experiment. *J. Mol. Biol.* 336:241–251.
12. Zagrovic, B., and V. S. Pande. 2003. Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem.* 24:1432–1436.
13. Bursulaya, B. D., and C. L. Brooks. 1999. The folding free energy surface of a three-stranded beta-sheet protein. *J. Am. Chem. Soc.* 121:9947–9951.
14. Garcia, A. E., and K. Y. Sanbonmatsu. 2002. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA*. 99:2782–2787.
15. Daggett, V., and M. Levitt. 1992. Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* 223:1121–1138.
16. Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA*. 100:13898–13903.
17. Cheung, M. S., J. M. Finke, B. Callahan, and J. N. Onuchic. 2003. Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B*. 107: 11193–11200.
18. Chan, H. S., and K. A. Dill. 1993. The protein folding problem. *Phys. Today*. 46:24–32.
19. Clementi, C., P. A. Jennings, and J. N. Onuchic. 2000. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc. Natl. Acad. Sci. USA*. 97:5871–5876.
20. Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* 83:3525–3532.
21. Shea, J. E., J. N. Onuchic, and C. L. Brooks. 1999. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA*. 96: 12512–12517.
22. Klimov, D. K., and D. Thirumalai. 2000. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci. USA*. 97:2544–2549.
23. Godzik, A., J. Skolnick, and A. Kolinski. 1992. Simulations of the folding pathway of triose phosphate isomerase-type alpha/beta barrel proteins. *Proc. Natl. Acad. Sci. USA*. 89:2629–2633.
24. Finke, J. M., M. S. Cheung, and J. N. Onuchic. 2004. A structural model of polyglutamine determined from a host-guest method combining experiments and landscape theory. *Biophys. J.* 87:1900–1918.
25. Benitez-Cardoza, C. G., A. Rojo-Dominguez, and A. Hernandez-Arana. 2001. Temperature-induced denaturation and renaturation of triosephosphate isomerase from Saccharomyces cerevisiae: evidence of dimerization coupled to refolding of the thermally unfolded protein. *Biochemistry*. 40:9049–9058.
26. Najera, H., M. Costas, and D. A. Fernandez-Velasco. 2003. Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme. *Biochem. J.* 370:785–792.
27. Silverman, J. A., and P. B. Harbury. 2002. The equilibrium unfolding pathway of a (beta/alpha)8 barrel. *J. Mol. Biol.* 324:1031–1040.
28. Chanez-Cardenas, M. E., D. A. Fernandez-Velasco, E. Vazquez-Contreras, R. Coria, G. Saab-Rincon, and R. Perez-Montfort. 2002. Unfolding of triosephosphate isomerase from Trypanosoma brucei: identification of intermediates and insight into the denaturation pathway using tryptophan mutants. *Arch. Biochem. Biophys.* 399:117–129.

29. Vadrevu, R., C. J. Falzone, and C. R. Matthews. 2003. Partial NMR assignments and secondary structure mapping of the isolated alpha subunit of Escherichia coli tryptophan synthase, a 29-kD TIM barrel protein. *Protein Sci.* 12:185–191.

30. Beasty, A. M., and C. R. Matthews. 1985. Characterization of an early intermediate in the folding of the alpha subunit of tryptophan synthase by hydrogen exchange measurement. *Biochemistry.* 24:3547–3553.

31. Forsyth, W. R., and C. R. Matthews. 2002. Folding mechanism of indole-3-glycerol phosphate synthase from Sulfolobus solfataricus: a test of the conservation of folding mechanisms hypothesis in (beta(alpha))(8) barrels. *J. Mol. Biol.* 320:1119–1133.

32. Jasanoff, A., B. Davis, and A. R. Fersht. 1994. Detection of an intermediate in the folding of the (beta alpha)8-barrel N-(5′-phosphoribosyl)anthranilate isomerase from Escherichia coli. *Biochemistry.* 33:6350–6355.

33. Soberon, X., P. Fuentes-Gallego, and G. Saab-Rincon. 2004. In vivo fragment complementation of a (beta/alpha)(8) barrel protein: generation of variability by recombination. *FEBS Lett.* 560:167–172.

34. Pan, H., and D. L. Smith. 2003. Quaternary structure of aldolase leads to differences in its folding and unfolding intermediates. *Biochemistry.* 42:5713–5721.

35. Deng, Y., and D. L. Smith. 1999. Rate and equilibrium constants for protein unfolding and refolding determined by hydrogen exchange-mass spectrometry. *Anal. Biochem.* 276:150–160.

36. Pan, H., A. S. Raza, and D. L. Smith. 2004. Equilibrium and kinetic folding of rabbit muscle triosephosphate isomerase by hydrogen exchange mass spectrometry. *J. Mol. Biol.* 336:1251–1263.

37. Rojsajjakul, T., P. Wintrode, R. Vadrevu, C. Robert Matthews, and D. L. Smith. 2004. Multi-state unfolding of the alpha subunit of tryptophan synthase, a TIM barrel protein: insights into the secondary structure of the stable equilibrium intermediates by hydrogen exchange mass spectrometry. *J. Mol. Biol.* 341:241–253.

38. Saab-Rincon, G., P. J. Gualfetti, and C. R. Matthews. 1996. Mutagenic and thermodynamic analyses of residual structure in the alpha-subunit of tryptophan synthase. *Biochemistry.* 35:1988–1994.

39. Gualfetti, P. J., O. Bilsel, and C. R. Matthews. 1999. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from Escherichia coli. *Protein Sci.* 8:1623–1635.

40. Zitzewitz, J. A., P. J. Gualfetti, I. A. Perkons, S. A. Wasta, and C. R. Matthews. 1999. Identifying the structural boundaries of independent folding domains in the alpha subunit of tryptophan synthase, a beta/alpha barrel protein. *Protein Sci.* 8:1200–1209.

41. Bilsel, O., J. A. Zitzewitz, K. E. Bowers, and C. R. Matthews. 1999. Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry.* 38:1018–1029.

42. Wu, Y., and C. R. Matthews. 2003. Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol.* 330:1131–1144.

43. Sanchez del Pino, M. M., and A. R. Fersht. 1997. Nonsequential unfolding of the alpha/beta barrel protein indole-3-glycerol-phosphate synthase. *Biochemistry.* 36:5560–5565.

44. Andreotti, G., M. V. Cubellis, M. D. Palo, D. Fessas, G. Sannia, and G. Marino. 1997. Stability of a thermophilic TIM-barrel enzyme: indole-3-glycerol phosphate synthase from the thermophilic archaeon Sulfolobus solfataricus. *Biochem. J.* 323:259–264.

45. Tsuji, T., B. A. Chrunyk, X. Chen, and C. R. Matthews. 1993. Mutagenic analysis of the interior packing of an alpha/beta barrel protein. Effects on the stabilities and rates of interconversion of the native and partially folded forms of the alpha subunit of tryptophan synthase. *Biochemistry.* 32:5566–5575.

46. Anderson, W. L., and D. B. Wetlaufer. 1976. The folding pathway of reduced lysozyme. *J. Biol. Chem.* 251:3147–3153.

47. Finke, J. M., and P. A. Jennings. 2001. Early aggregated states in the folding of interleukin-1β. *J. Biol. Phys.* 27:119–131.

48. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

49. Pearlman, D. A., D. A. Case, J. W. Caldwell, W. R. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. A. Kollman. 1995. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comput. Phys. Commun.* 91:1–41.

50. Berendsen, H. J. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

51. Hurle, M. R., N. B. Tweedy, and C. R. Matthews. 1986. Synergism in folding of a double mutant of the alpha subunit of tryptophan synthase. *Biochemistry.* 25:6356–6360.

52. Go, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.

53. Sobolev, V., A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman. 1999. Automated analysis of interatomic contacts in proteins. *Bioinformatics.* 15:327–332.

54. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.

55. Barshop, B. A., R. F. Wrenn, and C. Frieden. 1983. Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM—a flexible, portable system. *Anal. Biochem.* 130:134–145.

56. Socci, N. D., and J. N. Onuchic. 1994. Folding kinetics of proteinlike heteropolymers. *J. Chem. Phys.* 101:1519–1528.

57. Camacho, C. J., and D. Thirumalai. 1995. Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.* 92:1277–1281.

58. Kiefhaber, T. 1995. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA.* 92:9029–9033.

59. Kuwata, K., R. Shastry, H. Cheng, M. Hoshino, C. A. Batt, Y. Goto, and H. Roder. 2001. Structural and kinetic characterization of early folding events in beta-lactoglobulin. *Nat. Struct. Biol.* 8:151–155.